PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 9/42, 15/56

A1

(11) International Publication Number: WO 94/28117

(43) International Publication Date: 8 December 1994 (08.12.94)

(21) International Application Number: PCT/FI94/00234

(22) International Filing Date: 2 June 1994 (02.06.94)

(30) Priority Data:
932521 2 June 1993 (02.06.93) FI

(71) Applicant (*for all designated States except US*): OY ALKO AB [FI/FI]; Salmisaarenranta 7 H, FIN-00180 Helsinki (FI).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): SALOHEIMO, Anu [FI/FI]; Kyntäjäntie 18 B, FIN-00390 Helsinki (FI). SIIKA-AHO, Matti [FI/FI]; Susitie 26 E 33, FIN-00800 Helsinki (FI). PENTTILÄ, Merja [FI/FI]; Vähänivantie 9 A 6, FIN-00390 Helsinki (FI). LANTTO, Raija [FI/FI]; Jokitie 1, FIN-01800 Klaukkala (FI). MÄNTYLÄ, Arja [FI/FI]; Myllykalliontie 2 C 32, FIN-00200 Helsinki (FI). PALOHEIMO, Marja [FI/FI]; Oksasenkatu 7 A 24, FIN-00100 Helsinki (FI). SUOMINEN, Pirkko [FI/FI]; Kotitorpantie 11 as 4, FIN-00690 Helsinki (FI).

(74) Agents: LAINE, Seppo et al.; Seppo Laine Oy, Lönnrotinkatu 19 A, FIN-00120 Helsinki (FI).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KG, KP, KR, KZ, LK, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: NOVEL ENDOGLUCANASE ENZYME

(57) Abstract

The present invention concerns a novel enzyme, EGV, having endoglucanase activity. The molecular weight of the enzyme is about 20 to 25 kDA and it is isolated from the fungus *Trichoderma reesei*. The invention also relates to a DNA sequence coding for the novel enzyme as well as vectors, yeast strains and fungal strains containing the DNA sequence. Furthermore, the invention concerns a method for isolating the DNA sequence coding for the novel enzyme and for constructing yeast and fungal strains which are capable of expressing endoglucanase. The invention also provides an enzyme product having endoglucanase activity and methods for enzymatically modifying lignocellulosic materials, in particular for modification or degradation of cellulose and/or β-glucan.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

NOVEL ENDOGLUCANASE ENZYME

Field of the invention

The present invention concerns a novel enzyme having endoglucanase activity. The enzyme is isolated from the fungus *Trichoderma reesei*. The invention also relates to an isolated and purified DNA sequence coding for the novel enzyme as well as vectors, yeast strains and fungal strains containing the DNA sequence. Furthermore, the invention concerns a method for isolating the DNA sequence coding for the novel enzyme and for constructing fungal strains which are capable of expressing endoglucanase. The invention also provides an enzyme product having endoglucanase activity and methods for enzymatically modifying cellulosic/lignocellulosic materials, in particular for modification or degradation of cellulose and/or β-glucan.

Background of the invention

Many fungal species produce enzymes that degrade plant polymers into simple compounds like sugars. The fungus *Trichoderma reesei* is one of the most potent and most studied organisms degrading cellulose. It produces all the enzyme types needed for efficient break-down of crystalline cellulose, namely endo-1,4-ß-D-glucanases (EC 3.2.1.4), cellobiohydrolases (exo-1,4-ß-D-glucanases, EC 3.2.1.91) and 1,4-ß-D-glucosidases (EC 4.3.2.21). The number of enzymes belonging to each class is far from clear, but the existence of at least two cellobiohydrolases, CBHI and CBHII, and two endoglucanases, EGI and EGII (formerly EGIII,) has been confirmed by cloning of the corresponding genes (Shoemaker et al. 1983, Teeri et al. 1983, Penttilä et al. 1986, Chen et al. 1987, Teeri et al. 1987, van Arsdell et al. 1987, Saloheimo et al. 1988).

It is known in the art that the different types of cellulolytic enzymes mentioned above attack different parts of the cellulose molecule, and that cellulose hydrolyzation by a cellulase mixture is the result of synergy between its components. Therefore, if total hydrolysis of a cellulose substrate is aimed at, it is generally required that the cellulase mixture contain β-glucosidases, cellobiohydrolases as well as endoglucanases. As mentioned above,

Trichoderma reesei produces such enzyme mixtures.

It is also known that the cellulase enzymes belonging to the same class are mutually different as regards their activities towards some of the cellulosic substrates. For example, CBHII will catalyze hydrolysis of β-glucan whereas CBHI is inactive toward that substrate.

The cellulase enzymes usually consist functionally of two different parts, viz. a core and a tail, which are interconnected by an intermittent part (known as the linker). The active centre of the enzyme is located in the core. The function of the tail consists mainly of its capability to attach the enzyme to an insoluble substrate. Thus, if the tail is removed the activity of the enzyme toward macromolecular and crystalline substrates can be substantially decreased.

By way of a general definition, the name "endoglucanases" is assigned to enzymes that catalyze random hydrolysis of β-1-4 glycosidic bonds between glucose units of cellulose polymers. The two major *Trichoderma* endoglucanases, EGI and EGII, contain about 500 to 600 amino acids and their molecular weights are about 50 to 60 kDa. Also the cellobiohydrolases are similar in size. These kinds or rather bulky molecules may have difficulties in penetrating some fibrous substrates whose adjacent polysaccharide chains are aligned and located close to each other. Such substrates are represented by fibrous materials of great economic values, such as cellulose pulp. Therefore, endoglucanases of a low molecular weight have been of an increasing interest during the last years.

Håkansson et al. (1978) have purified a small endoglucanase from culture filtrates of *T. reesei*. This enzyme has a size of about 20 kDa, a neutral pI and, unlike the major cellulases, it does not contain carbohydrate moieties. Håkansson et al. found the enzyme to be present in the culture medium in very small amounts. Small endoglucanases of similar properties have also been isolated by Gong et al. (1979) and Ülker and Sprey (1990). However, although the molecular weight of the endoglucanase isolated by Håkansson and partially sequenced by Ståhlberg in 1991 is rather low, the molecular configuration of the enzyme is not advantageous as far as enzymatical applications are concerned. The molecule appears not to contain a linker domain and a cellulose binding domain (CBD) but only a core domain. The cellulose binding domain and a linker region, allowing for its flexible separation

from the catalytic core, are essential features of true cellulases capable of efficient attachment to the substrate.

The PCT Application No. PCT/US91/07276 discloses an endoglucanase enzyme, called EGIII, derived from *Trichoderma*. The molecular size of the EGIII is 23 to 28 kDa, its pH optimum is 5.5 to 6.0 and the pI 7.2 to 8.0. From the sequence data of EGIII, it is apparent that said enzyme is the same as the one isolated by Håkansson and sequenced by Ståhlberg and that it does not contain the linker and CBD domains.

Known in the art are also small endoglucanase enzymes isolated from other microorganisms. Thus, a gene coding for a polypeptide homologous to the short amino acid sequence available from the protein described above has been isolated from the fungus *Aspergillus aculeatus* (Ooi et al. 1990). The PCT Patent Application No. PCT/DK91/00123 describes an endoglucanase derived from the fungus *Humicola insolens*. The size of the polypeptide molecule is 43 kDa and its isoelectric point is 5.1. The use of the enzyme for treatment of cellulose-containing fabrics is suggested.

Nothing has so far been reported on the existance of a small size, true *Trichoderma* endoglucanase having cellulose binding regions. It has been a general conception that the cellulase system of *Trichoderma* consists of at least two CBH:s and two EGs and additionally of the EGIII which lacks a CBD.

Isolation and manipulation of the cellulase genes is very important for the various commercial uses of enzymes and of the organisms producing them. Isolation of hydrolase genes from eukaryotes has been a task demanding either extensive studies on the corresponding enzymes or the laborous differential hybridization protocols.

Summary of the invention

It is an object of the present invention to provide a novel endoglucanase enzyme of low molecular weight and having a suitable configuration for enzymatical applications.

This invention provides an endoglucanase enzyme derived from *Trichoderma reesei* which (in unglycosylated form) has a molecular weight of about 20 to 25 kDa and contains 242 amino acids (the mature protein contains less amino acids that that depending on the signal sequence cleavage site), some 70 % of which are located in the core region, whereas roughly one sixth of the amino acids is in the linker, taking an extended conformation, and one sixth in the CBD domain. This distribution of the amino acid residues within the molecule gives evidence of it having an elongated, "wormish" form in comparison to other cellulases, which facilitates penetration between adjacent molecules of fibrous cellulosic substrates. Being different in structure and activity, the enzyme complements the cellulolytic enzyme mixture acting in synergy, as the Examples below will show.

Another object of the invention is to provide a simple and rapid method for isolation of endoglucanase genes by function. In fact, the method described in more detail below, makes it possible to isolate any hydrolytic enzyme gene, such as genes coding for cellulases (for instance endoglucanases and cellobiohydrolases) and hemicellulases (for instance xylanases and mannanases), without previous knowledge of the corresponding proteins. In this connection it should be pointed out that before this invention there did not exist any data on protein level which would have suggested the existence of the novel endoglucanase described herein. This fact is already indicative of the unusual properties resulting in its disregard in the biochemical characterization of the cellulase mixture produced by *Trichoderma*.

According to the present method, an expression cDNA library is made from the organism of choice into a yeast expression vector. Yeast transformants are screened on plates containing the substrate of the desired activity. Using our earlier finding (Penttilä et al. 1987, 1988) that yeast produces and secretes the major cellulases of *T. reesei* in active form, the enzymatic activities can be visualized on substrate plates.

In the following description of the present invention, the novel gene coding for the novel endoglucanase enzyme is characterized as is its transfer into, and the expression thereof, in suitable hosts, such as fungi of the genus *Trichoderma*, in particular various *Trichoderma reesei* strains, and yeasts, such as *Saccharomyces cerevisiae*.

The invention also provides yeast and fungal strains transformed with the gene coding for the novel endoglucanase enzyme. Finally, applications of the enzyme are suggested.

Brief description of the drawings

Fig. 1 shows the nucleotide sequence of the gene *egl5* coding for the novel enzyme, EGV.

Fig. 2A shows the cellulose binding domains and Fig. 2B linker regions of EGV compared with the same domains and regions of the other *Trichoderma* cellulases. In Fig 2B, the serine and theronine residues have been boxed.

Fig. 3 shows the endoglucanase gene *egl5* integrated into plasmid pAJ401 resulting in plasmid pAS4.

Fig. 4 shows the endoglucanase gene *egl5* integrated into plasmid pMLO16del5 resulting in plasmid pAS16.

Fig. 5 shows the structure of plasmid pMLO16,

Fig. 6 shows the structure of plasmid pMLO16del5(11),

Figs. 7a to 7d depicts the construction of the *egl5* expression plasmid pALK956, Fig. 7a also indicating the structure of plasmid pAS13,

Fig. 8 indicates the relative activity of the novel endoglucanase enzyme as a function of the pH,

Fig. 9 shows the pH stability of the enzyme, and

Fig. 10 shows the introns and coding sequence of the *egl5* gene.

Detailed description of the invention

In the following description, the following abbreviations and definitions are used:

Abbreviations:

aa, amino acid(s); bp, base pair(s); CBD, cellulose-binding domain; CBH, cellobiohydrolase; *cbh*, gene coding for CBH; CMC, carboxymethyl cellulose; EG, endoglucanase; *egl*, gene coding for EG; HCA, hydrophobic cluster analysis; HEC, hydroxyethyl cellulose; kb, kilobase(s); kDa, kilo dalton(s); MUC, 4-methyl-umbelliferyl ß-D-cellobioside; MUL, 4-methyl-

umbelliferyl ß-D-lactoside; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; *PGK*, 3-phosphoglycerate kinase gene of *Saccharomyces cerevisiae*; pI, isoelectric point.

Definitions:

Within the scope of the present invention, the term "cellulase" is used as a collective term which encompasses enzymes catalyzing reactions which participate in the degradation of insoluble cellulose or cellulosic substrates to soluble carbohydrate. "Cellulase" is known in the art to refer to such a group of enzymes. As mentioned above, for hydrolysis of cellulose to glucose, three cellulase enzymes (three types of cellulase enzyme activity) are needed: randomly cleaving endoglucanases (1,4,-β-D-glucan glucanohydrolase, EC 3.2.1.4) which usually attack substituted soluble substrates; cellobiohydrolase (1,4-β-D-glucan cellobiohydrolase, EC 3.2.1.91) which is capable of degrading crystalline cellulose but has no activity towards derivatized cellulose and β-glucosidase (β-D-glucoside glycohydrolase, EC 3.2.1.21) which degrades cellobiose and cello-oligosaccharides to yield glucose. Each of the three main types of enzymes listed above occurs in multiple forms. For example, two immunologically distinctive cellobiohydrolases, CBHI and CBHII are known. In addition, at least two distinct endoglucanases are known. Synergistic action between some of these enzymes has been demonstrated. "Cellulase activity" is synonymous with cellulolytic activity.

Enzymes having "endoglucanase activity" are, within the scope of the present invention, enzymes which will catalyse the hydrolysis of internal β-1,4-linkages of cellulose.

By "enzyme preparation" is meant a composition containing enzymes which have been extracted from (either partially or completely purified from) the microorganisms (for instance the fungi) producing them. The term "enzyme preparation" is meant to include a composition comprising medium used to culture such microorganisms and any enzymes which the microorganisms have secreted into such medium during the culture.

"Culture medium" denotes a medium previously used to culture a fungi ("spent" culture medium), such culture medium containing enzymes which the fungi have secreted into the

medium during the culture. The culture medium can be used as such or as partially or completely purified, concentrated, dried or immobilized.

By "hybridization " are meant conditions, under which all the different forms of DNA sequences hybridize to the DNA sequence encoding for the *Trichoderma* enzyme having endoglucanase activity, the molecular weight of the unglycosylated form of said enzyme being about 20 to 25 kDa and containing 242 amino acids (the mature protein having less amino acids).

"Gene" denotes a DNA sequence containing a template for a RNA polymerase. RNA that codes for a protein is termed messenger RNA (mRNA).

It is well known that mutations occur in genes and can cause changes in the amino acid sequence of the encoded polypeptide. Changes can also be introduced by genetic engineering techniques. As used herein, the term *egl5* gene includes all DNA sequences homologous with the sequence herein disclosed for *egl5* and encoding polypeptides with the fuctional or structural properties of the about 20 to 25 kDa polypeptide. It is known in the art that cellulases lacking the linker and CBD regions still exhibit catalytic activity towards the β-1,4-glucosidic linkage, and thus a smaller core polypeptide is also included in the denotion of *egl5*. Sequences artificially derived from this gene but still encoding a polypeptide with the desired fuctional or structural properties are also included and encompassed by the expression "functional equivalents".

A cloning vehicle or a vector is a plasmid or phage DNA or other DNA sequence (such as a linear DNA) which provides an appropriate nucleic acid environment for the transfer of a gene of interest into a host cell. The cloning vehicles of the invention may be designed to replicate autonomously in prokaryotic and eukaryotic hosts. In *Trichoderma*, the cloning vehicles generally do not autonomously replicate and instead, merely provide a vehicle for the transport of the gene of interest into the *Trichoderma* host for subsequent insertion into the *Trichoderma* genome. The cloning vehicle may be further characterized by one or a small number of endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vehicle, and into

which DNA may be spliced in order to bring about replication and cloning of such DNA. The cloning vehicle may further contain a marker suitable for use in the identification of cells transformed with the cloning vehicle. Markers, for example, are tetracycline resistance or ampicillin resistance for *E. coli* and for example phleomycin resistance or acetamidase for *Trichoderma*. The word "vector" is sometimes used for "cloning vehicle. " Alternatively, such markers may be provided on a cloning vehicle which is separate from that supplying the gene of interest.

A vehicle or vector similar to a cloning vehicle but which is capable of expressing a gene of interest which has been cloned into it, after transformation into a desired host, is called an expression vector. In a preferred embodiment, such expression vehicle provides for an enhanced expression of a gene of interest which has been cloned into it, after transformation into a desired host.

The gene of interest which is provided to a fungal host as part of a cloning or expression vehicle integrates into the fungal chromosome. Sequences which derive from the cloning vehicle or expression vehicle may also be integrated with the gene of interest during the integration process.

The gene of interest may preferably be placed under the control of (i.e., operably linked to) certain control sequences such as promoter sequences provided by the vector (which integrate with the gene of interest). If desired, such control sequences may be provided by the fungal host's chromosome as a result of the locus of insertion.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains expression control sequences which contain transcriptional regulatory information and such sequences are "operably linked" to the nucleotide sequence which encodes the polypeptide.

Method of isolating genes

Bacterial cellulase genes have widely been isolated by transforming genomic libraries into *E. coli* and screening activities on cellulose-containing plates (reviewed by Béguin et al. 1987). This approach relies on the functionality of promoter sequences from other prokaryotes in *E. coli* and is not applicable to eukaryotes. Furthermore, eukaryotic genes, such as the *T. reesei* EGV described here, contain introns which cannot be excised in *E. coli* and thus disturb the reading frame. Moreover, the *Trichoderma* cellulases cannot generally be expressed in *E. coli* in active form even if expressed from cDNA coupled to bacterial expression signals. Traditionally fungal cellulase genes have been cloned using either differential hybridization, antibodies raised against the corresponding enzymes or hybridization with oligonucleotide probes based on the protein sequence of the enzymes (Béguin et al. 1987). All these methods are laborous and demand a lot of time and previous knowledge of the corresponding enzymes.

With the method according to the present invention, genes coding for new activities can be easily isolated without any previous knowledge of the protein. According to the invention, a fungal strain (e.g. *Trichoderma*) is cultivated on a culture medium which will induce enzyme production. Such culture medium typically contains cellulosic substrate, if endoglucanase production is aimed at. After cultivation, the mRNA of the strain is isolated and the corresponding cDNA is formed. cDNA made from the organism of interest is cloned into a yeast vector to construct an expression gene library in yeast, for instance *Saccharomyces cerevisiae*. The genes of the fungus are then expressed under any suitable promoter providing sufficient expression level, such as the yeast promoter *PGK*. The enzyme, e.g. endoglucanase, is extracellularly secreted and the colonies producing the desired enzymes, e.g. the endoglucanase, can be identified on the basis of their production of enzyme activity.

Screening can be effected with activity plate assays. Thus, according to one preferred embodiment of the present invention, the endoglucanase gene is isolated by plating the expression library onto plates containing barley ß-glucan as substrate. After growth the cells are washed away and the plates are stained with congo red to reveal the hydrolysis halos. Up to 50 % of the clones giving halos may contain endoglucanase. The genes coding for

different endoglucanases can be identified by analyzing the clones.

The expression gene library can also be constructed by using some other yeast promoter which will provide a weaker level of expression. If it is to be expected that the enzyme is deleterions to the yeast, the inducible GAL1 promoter would be recommendable. It is also possible to use the endoglucanase's own promoter and, for the purpose of isolating the genes, a chromosomal gene library can, in some cases, be used. The gene library can also be constructed in a single copy plasmid. Also any other yeast strain with established transformation procedures can be used as a host, because their secretion capabilities are usually even higher than that of *Saccharomyces*.

In summary, the invention comprises the steps of

- enriching the mRNA pool of a fungal strain, e.g. *Trichoderma*, producing endoglucanase activity in respect of the mRNA of the endoglucanase by cultivating the strain in conditions which will induce the endoglucanase production of said strain,
- isolating mRNA from the strain,
- preparing cDNA corresponding to the isolated mRNA,
- placing the cDNA thus obtained in a vector under the control of a suitable promoter,
- transforming the recombinant plasmids into a yeast strain which naturally does not produce significant amounts of the endoglucanase in order to provide an expression library,
- cultivating the yeast clones thus obtained on a cultivation medium in order to express the expression library in the yeast,
- separating the yeast clones producing endoglucanase from the other yeast clones,
- isolating the plasmid-DNA of said separated yeast clones, and,
- if desired, sequencing the DNA in order to determine the DNA sequence coding for the endoglucanase.

Endoglucanase V and the gene *egl5*

The gene *egl5* isolated according to the above method was sequenced according to conventional methods. The DNA sequence of *egl5* is shown in Figure 1 and also indicated in SEQ ID NO. 1.

The gene *egl5* codes for a previously unknown protein of 242 amino acids, the amino acid sequence of which is depicted in SEQ ID NO. 2. Interestingly, this protein contains the two conservative domains found in all *Trichoderma* cellulases, namely the cellulose-binding domain (CBD) and the linker region that connects the CBD to the catalytic core domain. The approximate regions comprising these domains are indicated in Figure 1, the linker region being the part of the sequence marked with the letter B, whereas the cellulose binding domain is marked with the letter A. The putative N-glycosylation site is marked with an asterisk. At the beginning of the protein a 17 amino acid long signal sequence (Met-Lys-Ala-Thr-Leu-Val-Leu-Gly-Ser-Leu-Ile-Val-Gly-Ala-Val-Ser-Ala), which is underlined in Figure 1, can be predicted. If the signal sequence cleavage occurs at this position, the mature protein consists of 225 amino acids and has a calculated molecular weight of 22.799 KDa.

The core of the endoglucanase is separately depicted in SEQ ID NO. 3. It would appear that the core of the novel endoglucanase is primarily responsible for the cellulolytic activity of the novel enzyme. Thus, it is conceived that an endoglucanase enzyme product may in principle comprise the polypeptide of the core domain only. However, the surprising enzymatic properties described below are probably attributable to a combination of the above three regions and domains, and they will therefore best be obtained if the protein comprises all three parts.

It is believed that the predicted 17 aa signal peptide indicated in Figure 1 can be substituted by another suitable signal peptide possibly of a different length. Such a signal sequence should typically comprise a positively charged amino acid at the beginning followed by a stretch of hydrophobic amino acids. Depending on the signal sequence cleavage site *in vivo* and the possible proteolytic processing occurring frequently in cellulases, the molecular weight of the active polypeptide may vary somewhat and the novel endoglucanase is

therefore referred to as having a molecular weight in unglycosylated form of about 20 to 25 kDa.

In its O- and N-glycosylated form the enzyme can be significantly bigger having apparent molecular weights of 35 kDa or even much higher when produced in the yeast *Saccharomyces*. Furthermore, cellulases frequently undergo drastic proteolytic cleavage which removes the CBD (and linker) regions and consequently the size of EGV in fungal culture medium can be even about 115 kDa in unglycosylated form.

While modelling protein conformations from first principles is not possible, the high sequence similarity between fungal CBDs warrants the construction of a homology model (Sali *et al.*, 1990). The feasibility of modelling side chain conformations has been demonstrated in similar cases (Blundell *et al.*, 1988, Heiner *et al.*, 1993). Modelling of the EGV CBD revealed some interesting differences compared to the known structure of the CBHI CBD. This wedge-shaped domain seems less sharp in EGV and there are some differences in main chain and side chain conformations and in hydrophobic properties in areas known to be important for binding of the CBHI CBD onto the cellulose surface or for the full activity of the CBHI enzyme against crystalline cellulose. Preliminary binding data indicate that the EGV CBD is able to bind to cellulose.

The protein belongs to a new family K of cellulases together with the endoglucanase B of *Pseudomonas fluorescens* and the endoglucanase V of *Humicola insolens* as studied by hydrophobic cluster analysis by Henrissat and Bairoch (1993). This strongly suggests that EGV is structurally different from all *Trichoderma* cellulases characterized so far. Based on this, it would also appear that there are catalytic differences between the present enzyme and the other cellulases. The fact that EGV is a true endoglucanase was confirmed by ^{1}H-NMR spectroscopy, which showed that the internal β-1,4-linkages were hydrolysed by EGV when barley β-glucan (a soluble glucose polymer containing β-1,4- and β-1,3-linkages) was used as substrate.

Thus, for instance, as evidenced by Example 11 below, the novel endoglucanase appears to work synergetically with the known endoglucanase EGII on hydroxyethyl cellulose.

Expression of the gene *egl5*

Once the vector or DNA sequence containing the construct(s) is prepared for expression, the DNA construct(s) is introduced into an appropriate host cell by any of a variety of suitable means, including transformation as described above. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of transformed cells. Expression of the cloned gene sequence(s) results in the production of the desired protein, or in the production of a fragment of this protein. This expression can take place in a continuous manner in the transformed cells, or in a controlled manner.

Expression of the gene can be obtained in any fungus with developed transformation and expression methods.

Trichoderma is an especially useful and practical host for the synthesis of the enzyme preparations of the invention because *Trichoderma* is capable of secreting protein at large amounts, for example, concentrations as much as 40 g/L culture fluid have been reported; the homologous *Trichoderma cbh1* promoter provides a very convenient promoter for expression of genes-of-interest because it is a strong, single copy promoter which normally directs the synthesis of up to 60 % of the secreted protein from the *Trichoderma* host; the transformation system is highly versatile and can be adapted for any gene of interest; the *Trichoderma* host provides an "animal cell type" high mannose glycosylation pattern; and culture of *Trichoderma* is supported by previous extensive experience in industrial scale fermentation techniques. In addition, several promoters active on glucose medium can be used, which enable the production of the enzyme essentially free from other cellulases.

Expression of the protein in the *Trichoderma* hosts requires the use of regulatory regions functional in such hosts. A wide variety of transcriptional and translational regulatory sequences can be employed, since *Trichoderma* generally recognize eukaryotic host transcriptional controls, such as, for example, those of other filamentous fungi. Such control regions may or may not provide an initiator methionine (AUG) codon, depending on whether the cloned sequence contains such a methionine. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis in the host cell.

According to the invention the DNA sequence encoding EGV can be transformed into *Trichoderma* and expressed, for example under the strong *cbh1* promoter, as described in EP-A 244,234 and US 5,298,405, or other promoter functional in *Trichoderma*. The DNA sequence coding for EGV can be integrated into the general expression vector pAHM110. The transformation can be done as a cotransformation using two circular plasmids, the selection marker being located in one of the plasmids and the DNA sequence encoding *egl5* in the other, or the selection marker and the DNA sequence encoding the *egl5* can be located in the same plasmid, or linear fragments can be used in the transformation. Possible selection markers are, for instance, *trpC* or *argB* from *Aspergillus nidulans* or argB or pyr4 from *T. reesei* or *amdS* from *A. nidulans* or *trp1* from *Neurospora crassa* or phleomycine or hygromycine resistance markers from bacterial origin (EP-A 244,234, US 5,298,405, and EP-B 539,395 and Ulhoa et al., 1992, Transformation of *Trichoderma* species with dominant selectable markers, Curr. Genet 21:23-26) or other selection marker shown to function in *Trichoderma* in future (Karhunen et al. 1993, High frequency one-step gene replacement in *Trichoderma reesei* I, Endoglucanase I overproduction, MGG, 241: 515-522, and Suominen et al., 1993, High frequency one-step gene replacement in *Trichoderma reesei* II, Effects of deletions of individual cellulase genes. MGG, 241: 523-530.

To construct a *Trichoderma* strain producing endoglucanase V as the main cellulolytic enzyme it is possible to construct *Trichoderma* strains that do not produce the endoglucanases I and II or all other cellulolytic enzymes: endoglucanase I and II and cellobiohydrolase I and II. The desired cellulolytic genes can be made deficient (EP-A 244,234, US 5,298,405, Karhunen et al. (1993) and Suominen et al. 1993). If genes are expressed under the *cbh1* promoter the expression is repressed by glucose and thus the strains must be grown on cellulose-containing medium.

Alternatively, it is possible to construct *Trichoderma* strains expressing EGV under glucose promoter. This means that the *Trichoderma* strains expressing EGV can be grown on glucose containing medium. Possible glucose promoters are, for example, glucose derepressed *cbh1* promoter of the plasmid pMLO16del5(11) (*et al.*, 1992) and the promoter of the cDNA1 gene (Nakari *et al.*, 1992) or other glucose promoters.

According to the invention, there is also provided a method for producing in fungal and yeast hosts, such as the yeast *Saccharomyces* and filamentous fungi, such as *Trichoderma*, an enzyme preparation having an endoglucanase activity stemming from an endoglucanase enzyme, the molecular weight of which (in unglycosylated form) is 20 to 25 kDa.

Further, if desired activities are present in more than one recombinant host, such preparations can be isolated from the appropriate hosts and combined prior to use in the method of the invention.

Enzyme preparations

To obtain the enzyme preparations of the invention, containing elevated levels of the EGV, the recombinant hosts described above having the desired properties (that is, hosts capable of expressing the novel endoglucanase enzyme) are cultivated under suitable conditions (cf. above), the desired enzymes are secreted from the host into the culture medium, and the enzyme preparation is recovered from said culture medium by methods known in the art.

As mentioned above, the enzyme preparation can be produced by cultivating the fungal strain in conditions where the regulatory regions directing endoglucanase expression are operating, such as on a glucose-containing medium if the yeast *PGK* or *Trichoderma* glucose promoters are used. Thus, if endoglucanase V is expressed under glucose promoter, the *Trichoderma* strains can be grown on, e.g., glucose minimal medium (Penttilä et al, 1987) or other glucose containing medium, for example Bacto-Peptone 5 g/l, Yeast extract 1 g/l, KH_2PO_4 4 g/l, $(NH_4)_2SO_4$ 4 g/l, $MgSO_4$ 0,5 g/l, $CaCl_2$ 0,5 g/l and trace element $FeSO_4$"$7H_2O$ 5 mg/l, $MnSO_4.H_2O$ 1,6 mg/l, $ZnSO_4.7H_2O$ 1,4 mg/l and $CoCl_2.6H_2O$ 3.7 mg/l, pH 5.0 - 6.0.

The enzyme can be produced also in other conditions, such as on Solca floc cellulose, if the *Trichoderma cbh1* promoter is used, or on a galactose-containing medium, if the yeast galactose-inducible promoter is used. The cellulose-containing cultivation medium may, for instance, comprise, 6 % Solca floc cellulose (BW40, James River Corporation, Hackensack, NJ), 3 % distiller's spent grain, 0.5 % KH_2PO_4, 0.5 % $(NH4)_2SO_4$, and 0.1 % struktol as an antifoaming agent (struktol SB 2023, Schill & Seilacher, Hamburg, FRG). *Trichoderma*

strains are sensitive to glucose repression and require an inducer (cellulose, lactose or sophorose). The pH should preferably be kept at approximately pH 5 to 6 by the addition of phosphoric acid or ammonia and the temperature at 30 °C during the cultivation.

The enzyme preparation is recovered from the culture medium by using methods well known in the art. However, the enzyme preparations of the invention may be utilized directly from the culture medium with no further purification. If desired, such preparations may be lyophilized, immobilized or the enzymatic activity otherwise concentrated and/or stabilized for storage.

If desired, the expressed endoglucanase protein may be further purified in accordance with conventional conditions, such as extraction, precipitation, chromatography, affinity chromatography, electrophoresis, or the like.

Applications of the novel enzyme

The catalytic core of the novel enzyme is the smallest of fungal or bacterial cellulases characterized. Therefore the enzyme and the enzyme preparations according to the invention have application in the treatment of pulp and paper and in the textile industry. Furthermore, the enzyme can be used in the fodder industry. The properties of the novel endoglucanase are unexpected for a endoglucanase on basis of general knowledge.

Being a β-glucanase, the novel enzyme can be used for hydrolyzation of the β-glucan of barley. As a result, the viscosity of the fodder is lowered and the nutritional value of the fodder is improved.

As evidenced in Example 8, the pH optimum of the enzyme is higher than those of the other endoglucanases produced by strains of the species *Trichoderma*. This favorable pH range can be utilized in many ways. One preferred application is for removing colour from denim jeans; in acidic pH, reabsorption of the colour occurs, but at neutral pH there is much less reabsorption. Another preferred embodiment comprises deinking. Normally, the pH of a slurry of water and newsprint is about 5.5 to 6.0 and therefore the novel enzyme can be used

without any need for adjustment of the pH. On the other hand, coated paper contains fillers and pigments which will raise the pH of an aqueous paper slurry formed therefrom. If the pH of the slurry is lowered by adding mineral acid, at least some of the suspended or dissolved fillers and pigments may precipitate, e.g. in the form of calcium sulphate.

The small size and the advantageous pH range of the novel enzyme make it possible to use it for treating recycled fibre in order to improve the technical properties thereof. The enzyme is also applicable for improving pulp drainage.

The invention is described in more detail with the aid of the following non-limiting examples.

In the examples, the following strains and vectors were employed: *E. coli* strains PLK-F', pBluescript SK⁻, and XL-1-Blue (Stratagene) were used as hosts for plasmids and PLK-F' as a host for the cDNA library. The following plasmids were used: pAS11, pAS13, pALK487 and pALK183. The *T. reesei* strain QM9414 was used as a source of RNA for cDNA preparation and Northern analysis. *T. reesei* ALKO2221 and ALKO3524 were used as hosts for EGV expression. *S. cerevisiae* strain DBY746 (α his3 1 leu2-3 leu2-112 ura3-52 trp1-289 cyh^r cir$^+$) was used as a host for the expression library. Strain MD40-4c (α ura2 trp1 leu2-3 leu2-112 his3-11 his3-15) was used as a host for the plasmids pMP311, pMS3, pMP11 and pMP29 carrying *egl1*, *egl2*, *cbh1* and *cbh2* genes of *T. reesei*, respectively (Penttilä et al. 1987, 1988). The yeast expression vector pFL60 (Minet and Lacroute 1990) containing the constitutive yeast *PGK* promoter and terminator, *URA3* marker gene and the 2 micron plasmid replication origin was kindly provided by Dr. M. Minet, Centre de Génétique Moléculaire, C.N.R.S., France.

Example 1

Isolation of endoglucanase gene by expression in yeast and hydrolytic properties of the yeast

T. reesei strain QM9414 was cultivated in a 10 liter fermentor at 28 °C and pH 4.0 for 42 hours. The cultivation medium used to induce hydrolytic enzyme production contained 2 % Solka floc cellulose, 1 % distiller's spent grain, 0.2 % Locust bean gum -galactomannan (Serva), 0.5 % KH_2PO_4 and 0.5 % $(NH_4)_2SO_4$. After 42 hours of growth, lactose (Sigma), Birke 150 acetylglucuronoxylan and Oat spelt arabinoxylan were added in an amount of 0.1 % each and the cultivation was continued for further 24 hours.

Total RNA from the *T. reesei* strain was isolated as described by Chirgwin et al. (1979), and the poly(A)$^+$ fraction was separated by chromatography through oligo(dT)-cellulose (BRL). cDNA, synthesized by the ZAP-cDNA synthesis kit (Stratagene), was ligated to the *Eco*RI-*Xho*I cut plasmid pAJ401. Plasmid pAJ401 was derived from plasmid pFL60 (Minet and Lacroute 1990) by changing the two cloning sites *Eco*RI and *Xho*I between the yeast *PGK* promoter and terminator into the reverse orientation. Transformation of *E. coli* strain PLK-F' by electroporation (Bio-Rad) according to the manufacturer's instructions yielded a library of 3.5×10^4 independent clones. Plasmids were isolated from the pool of *E. coli* transformants and transformed into *S. cerevisiae* strain DBY746 by electroporation (Bio-Rad) according to the manufacturer's instructions. Electroporation with 7 µg of plasmid DNA yielded a library of 8×10^4 yeast transformants.

1.2×10^5 yeast cells were plated on barley ß-glucan-containing plates to a density of 2000 colonies / 85 mm plate and grown at 30 °C for 3 days. Colonies were replicated and the original plates stained with Congo Red. Unstained areas around yeast colonies indicate hydrolysis of the substrate to oligosaccharides. Colonies showing activity were picked up from the replica plates and purified on new activity plates. Plasmids were recovered from the purified clones and analysed by restriction enzyme digestions. 20 clones gave a similar pattern of bands which was clearly different from the earlier isolated cellulase genes of *T. reesei*.

Transformation of the plasmids back to yeast confirmed that the activities were caused by the cDNA inserts. One of these plasmids, pAS4 (cf. Figure 3), was studied further. The insert in the pAS4 plasmid was named *egl5* and the corresponding protein EGV.

egl5 cDNA was sequenced from both strands of the original pAS4 plasmid using the Sanger dideoxynucleotide method, T7 DNA polymerase (Pharmacia) and sequence specific primers.

The sequence obtained is shown in SEQ ID NO. 1.

The chromosomal *egl5* gene was isolated from a *T. reesei* cosmic library (Mäntylä, A. et al. Curr. Genet. 1992, 21 471-477) by using the *egl5* cDNA as a probe. About 6 kb *Hind*III fragment was subcloned to pBluescript SK$^-$, resulting in plasmid pAS13 (Fig. 7a). The introns and coding sequence of *egl5* gene are shown in Figure 10 (SEQ ID NO. 11).

The activities of the yeast strain DBY746 carrying the pAS4 plasmid were studied by plate assays and they were compared with the activities of the yeast strains producing CBHI, CBHII, EGI and EGII.

Hydrolytic activities produced by recombinant yeast cells were detected on SD plates containing 0.1 % barley β-glucan (β-D-1,3-1,4-glucan, viscosity 20-30 c.s.; Biocon, UK, Sherman 1991) or hydroxyethyl cellulose (HEC, Fluka, Switzerland, product 54290). After growth the plates were stained with Congo Red (Merck) as described by Penttilä *et al.* (1987) to reveal the hydrolysis halos. Xylanase activity plates containing 0.2 % of a Remazol Brilliant Blue-dyed derivative of xylan (RBB-xylan, Sigma) needed no further treatment. Activities against synthetic substrates, 4-methylumbelliferyl β-D-cellobioside (MUC; Koch-Light, UK) or 4-methylumbelliferyl β-D-lactoside (MUL; Lambda Probes & Diagnostics, Austria) were detected as described by Penttilä *et al.* (1987).

The EGV protein showed a clear activity against β-glucan but the activity was lower than the activities of the strains producing EGI, CBHII or EGII (Table). However, the expression levels and the secretion efficiencies of foreign proteins in yeast may vary and thus it is not possible to draw any definite conclusions concerning the level of enzyme activity. Also, the

pH on the plates is not optimal for EGV function. EGV shows some activity against hydroxyethyl cellulose (HEC) in plate assays. No activity was detected on plate assays towards RBB-xylan or the small synthetic substrates, methylumbelliferyl cellobioside (MUC) or methylumbelliferyl lactoside (MUL).

Table 1 Hydrolytic activities of the yeast strains carrying the cellulase genes of *Trichoderma reesei*. The extent of hydrolysis of the substrate was estimated visually and is indicated by +.

	EGV	EGI	EGII	CBHI	CBHII
β-glucan	++	+++++	+++	-	++++
HEC	+	++++	++	-	-
MUL	-	+++	-	++	-
MUC	-	+++	-	+	-
RBB-xylan	-	+++	-	-	-

Example 2

Construction of endoglucanase expression vectors with truncated fragments of the *cbh1*-promoter

The vector pMLO16 (Figure 5) contains a 2.3 kb *cbh1* promoter fragment (SEQ ID 4) starting at 5' end from the *Eco*RI site, isolated from chromosomal gene bank of *Trichoderma reesei* (Teeri et al, 1983), a 3.1 kb *Bam*HI fragment of the *lac*Z gene from plasmid pAN924-21 (van Gorcom et al., 1985) and a 1.6 kb cbh1 terminator (SEQ ID 5) starting from 84 bp upstream from the translation stop codon and extending to a *Bam*HI site at the 3' end (Shoemaker et al. 1983; Teeri et al., 1983). These pieces were linked to a 2.3 kb long *Eco*RI-*Pvu*II region of pBR322 (Sutcliffe, J.G., 1979) generating junctions as shown in Figure 5. The exact in frame joint between the 2.3 kb *cbh1* promoter and the 3.1 kb *lac*Z gene was constructed by using an oligo depicted in Figure 5. A polylinker shown in Figure 5 was

cloned into the single internal *Xba*I site in the *chb1* promoter for the purpose of promoter deletions. A short *Sal*I linker shown in Figure 5 was cloned into the joint between the pBR322 and *cbh1* promoter fragments so that the expression cassette can be released from the vector by restriction digestion with *Sal*I and *Sph*I. Progressive unidirectional deletions were introduced to the *cbh1* promoter by cutting the vector with *Kpn*I and *Xho*I and using the Erase-A-Base System (Promega, Madison, USA) accordign to manufacturer's instructions. Plasmids obtained from different deletion time points were transformed into the *E. coli* strain DH5α (BRL) by the method described in (Hanahan D, 1983) and the deletion end points were sequenced by using standard methods.

Example 3

Construction of vectors for expression of EGV in *Trichoderma* in glucose-containing medium

In order to produce EGV protein in *Trichoderma reesei* QM9414 strain essentially free of other cellulases in a medium containing glucose, the plasmid pAS16 (Fig. 4) was constructed. There, the *egl5* cDNA was cloned under the truncated, glucose derepressed *cbh1* promoter of the plasmid pMLO16del5(11), generated as explained in Example 2. The plasmid contained a 1110 bp deletion in the *cbh1* promoter beginning from the promoter internal polylinker and ending 385 bp before the translation initiation site (Fig. 5). The sequence of this truncated promoter is provided as SEQ ID NO. 6. Plasmid pMLO16del5(11) was digested with the restriction enzymes *Ksp*I and *Sma*I. The vector part containing the glucose-derepressed *cbh1* promoter, the *cbh1* terminator and the pBR322 sequence was blunt-ended with the Mung bean nuclease, dephosphorylated with Calf intestin alkaline phosphatase and ligated to the *egl5* cDNA fragment.

The yeast expression plasmid pAS4 was digested with *Eco*RI and partially with *Xho*I to isolate the full-length *egl5* cDNA. The ends of the cDNA were filled-in with the Klenow polymerase enzyme and the fragment was ligated into the *Sma*I-cleaved vector pSP73 (Promega). The resulting plasmid pAS11 was digested with *Eco*RI and *Xba*I, filled-in with the Klenow polymerase and ligated to the vector part of the expression vector pMLO16del5(11).

Twenty micrograms of the pAS16 plasmid were digested with *Eco*RI and *Sph*I, phenol-extracted, precipitated and transformed into *Trichoderma reesei* QM9414 together with three micrograms of the plasmid p3SR2 (Hynes *et al.*, 1983) containing the acetamidase gene according to Penttilä *et al.*, (1987).

The promoter of the cDNA1 gene (Nakari *et al.*, 1992) was also used to direct the synthesis of the EGV protein on glucose-containing medium.

The promoter of the cDNA1 gene was cloned from the chromosomal DNA by PCR using the 5'primer GGT CTG AAG GAC GTG GAA TGA TGG (SEQ ID NO. 7) and the 3'primer GAT GCA TCG ATC GTC CGC GGG TTG AGA GAA GTT GTT GGA TTG ATC AAA AAG (SEQ ID NO. 8). The underlined ATCGAT in the 3'primer is a *Cla*I site and the CCGCGG a *Ksp*I site.

The *egl5* cDNA and the *cbh1* terminator were cloned as one fragment from the plasmid pAS16 by PCR using the 5'primer GAG AGA CCG CGG TGA TCT TCC ATC TCG TGT CTT GCT TGT AAC (SEQ ID NO. 9) and the 3'primer ATC GTG GAT CCA TTA TTA ACA CTT CGG TGG (SEQ ID NO. 10). The underlined CCGCGG in the 5'primer is a *Ksp*I site.

Eight micrograms of both of the fragments were digested with the *Ksp*I enzyme, purified from agarose gel and ligated. The ligation mixture was extracted with phenol, precipitated and used instead of a plasmid in the *Trichoderma* transformation together with three micrograms of the p3SR2 plasmid.

The Amd$^+$ transformants from the pAS16 transformation were streaked twice onto plates containing acetamide (Penttilä *et al.*, 1987), and then cultivated on Potato Dextrose Agar plates (Difco) from which spore suspensions were made. EGV production was tested from 50 ml shake flask cultures carried out in minimal medium according to Penttilä *et al.*, (1987) except that the amount of glucose was 4 %, KH_2PO_4 3 %, K_2PO_4 0.8 %, $(NH_4)_2SO_4$ 0.2 % and the medium was supplemented with 0.2 % peptone. Glucose was added as 15 % solution when necessary to keep the level above 1 % during the whole four days of the cultivation. The culture supernatants of 55 transformants were analyzed for activity against barley β-glucan by the DNS-method (Zurbriggen *et al.*, 1990).

The spore suspensions of the three best EGV-producing clones (numbers 101, 79 and 19) were purified to single spore cultures on Potato Dextrose Agar plates. EGV production was analyzed again from these purified clones as described above. The best producing transformant 101c was analysed by Southern blotting using conventional methods and the presence of the expression casette in the genomic DNA was confirmed. Northern analysis showed that the *egl5* gene was expressed from the constructs on glucose medium.

Example 4

Construction of EGV expression plasmid pALK956

The expression plasmid pALK956 (Figs. 7d) contains:

1) *T. reesei egl5* gene fused to the *cbh1* promoter. A fragment containing the *cbh1* terminator was included after *egl5* to ensure stop in the transcription.
2) *E. coli hph* (hygromycin B phosphotransferase; Gritz and Davies, 1983) as a marker gene for transformation. The gene was expressed from the *T. reesei pki* (pyruvate kinase; Schindler *et al.*, 1993) promoter.
3) Elongated *cbh1* terminator as a flanking region to ensure stop in *pki* transcription and to target the expression cassette, together with the *cbh1* promoter fragment, to the *cbh1* locus.

The construction of pALK956 is shown in detail in Figs. 7a - 7d. For the construction, the plasmids pAS11, pAS13, pALK487 and pALK183 were used. The plasmid pAS11 contains the *egl5* cDNA (Fig. 1) and pAS13 contains the chromosomal *egl5* gene (Fig. 10). The plasmid pALK487 contains the *T. reesei cbh1* promoter (the 2.2 kb *Stu*I - *Sac*II fragment originally from the plasmid pAMH110; Nevalainen *et al.*, 1991) and *cbh1* terminator (the 0.7 kb *Ava*II fragment starting 113 bp before the stop codon of the *cbh1* gene; for the *cbh1* sequence, see Shoemaker *et al.*, 1983). The plasmid pALK183 contains *hph* gene under the control of the *pki* promoter. It was constructed from pRLM$_{ex}$30 (Mach *et al.*, 1994) by changing the *cbh2* terminator to 1.6 kb *cbh1* elongated terminator (*Ava*II - *Bam*HI fragment).

The exact fusion of the *egl5* gene to the *cbh1* promoter was done by PCR. The 5'-primer

contained the last 26 nucleotides of the *cbh1* promoter including the *Sac*II site and the first 18 nucleotides of the coding sequence of *egl5* (5'-CAATAGTCAACCGCGGACTGCGCATCAT-GAAGGCAACTCTGGTT; the *Sac*II site is underlined, *egl5* sequence is bolded). The 3'-primer contained 21 nucleotides of *egl5* sequence, including the *Bam*HI site about 0.7 kb from the beginning of *egl5* sequence (5' -GGGCGTGGGATCCGTCTCTTG; the *Bam*HI site is underlined). The plasmid pAS13 was used as a template in the PCR reaction.

The 0.7 kb PCR fragment (filled in with DNA polymerase I Klenow fragment and cut with *Bam*HI), containing the exact link between the *cbh1* promoter and the *egl5* gene, was ligated to *Pvu*II - *Bam*HI digested pAS11 to obtain pALK951. The fusion and the PCR fragment were sequenced to ensure that no mistakes had occurred in the PCR amplification. Plasmid pALK955, containing the fusion of the *egl5* to the *cbh1* promoter, was obtained by ligating *Eco*RI/Klenow - *Sac*II fragment from pALK951 between the *cbh1* promoter and terminator in the plasmid pALK487 (*Bam*HI/Klenow - *Sac*II). The *hph* marker gene (under the control of the *pki* promoter) and the *cbh1* 3'-flanking region (elongated terminator) were ligated to *Stu*I cut pALK955 from pALK952 (*Xho*I - *Hind*III fragment / Klenow) to construct pALK956.

The plasmid pALK952 was constructed from pALK183 by shortening the *pki* promoter compared to the promoter used in pALK183 and pRLM$_{ex}$30 (*Not*I/partial - *Xho*I, Klenow).

The 7.4 kb expression cassette from pALK956 can be removed with *Not*I digestion.

Thus, in summary, in the expression plasmid pALK956, the *egl5* gene is fused to the *cbh1* promoter. The *E. coli hph* (hygromycin B phosphotransferase) gene is used as a marker for the transformations. The *cbh1* 3'-flanking region (elongated terminator) is included to ensure stop in the *pki* transcription and to target the expression cassette, together with the promoter fragment, to the *cbh1* locus.

Example 5

Expression of EGV under the *cbh1* promoter in cellulase-inducing medium

The EGV expression plasmid, pALK956, was digested with *Not*I, and the 7.4 kb fragment

was purified from agarose gel. 2-3 µg of the linear fragment was transformed into *T. reesei* strains ALKO2221 and ALKO3524 according to Penttilä et al. (1987) with the modifications described in Karhunen et al. (1993). ALKO2221 is a low protease mutant from *T. reesei* VTT-D-79125 (Bailey and Nevalainen, 1981), prepared in our laboratory (A. Mäntylä). ALKO3524 is a strain derived from VTT-D-79125, where the *cbh2, egl2* and *egl1* genes have been deleted using the *A. nidulans trpC* (Yelton et al., 1984), *A. nidulans amdS* (Kelly and Hynes, 1985) and *Streptoalloteichus hindustanus phleo*^{r} (Mattern et al., 1987) marker genes, respectively. The method of one-step gene replacement with a linear fragment and flanking regions of the corresponding cellulase locus is described in Suominen et al. (1993).

HygB+ transformants were selected on plates containing *T. reesei* minimal medium (Penttilä et al., 1987) with 100 µg hygromycin/ml. Transformants were purified by single spore selection on selective medium and then cultivated on Potato Dextrose Agar. Purified transformants were grown in shake flasks in a medium containing 4 % whey, 1.5 % complex nitrogen source derived from grain, 1.5 % KH_2PO_4 and 0.5 % $(NH_4)_2SO_4$. Cultures were maintained at 30 °C and 250 rpm for 7 days. The culture supernatants were analyzed for activity against barley β-glucan at pH 6.3 by the DNS-method (Zurbriggen et al., 1990). Soluble protein was assayed by the method of Lowry et al. (1951) using bovine serum albumin as standard. The detection of the 67 kDa CBHI protein was done in SDS-PAGE followed by Coomassie Brilliant Blue staining. The results from the best EGV transformants and the corresponding host strains are shown in Table 2. In the EGV-transformants the β-glucanase activity measured at the optimum pH of EGV was enhanced about twofold.

Table 2. Expression of EGV under the *cbhI* promoter in cellulase-inducing medium

Strain	Protein (mg/ml)	β-glucanase activity (BU/ml)	CBHI (+/-)
ALKO2221	7.1	1945	+
EGV/ALKO2221/11	5.2	3262	+
EGV/ALKO2221/47	5.5	3236	-
EGV/ALKO2221/31	3.5	3757	-
EGV/ALKO2221/68	3.9	3479	+
ALKO3524	9.3	3338	+
EGV/ALKO3524/27	5.3	6770	+
EGV/ALKO3524/28	7.6	7588	+
EGV/ALKO3524/31	7.0	6622	+

Example 6.

Enzyme preparation containing EGV protein obtained from yeast where egl5 gene was expressed.

Saccharomyces cerevisiae DBY 746 containing the pAS4 plasmid was grown in a bioreactor (Chemap LF 20, working volume 16 l) on a standard YPD medium. The inoculum (5 times 200 ml) was grown in shake flasks in selective synthetic complete medium without uracil. Cultivation conditions were: temperature 30 °C, pH controlled between 5.2 and 5.9, aeration about 15 l min^{-1} and cultivation time 45 h. The yeast cells were separated from the medium by centrifugation and the culture supernatant was concentrated 4-fold by ultrafiltration (PCI ES 625 membranes).

The enzymatic activity in the concentrate was assayed by standard methods using appropriate incubation times for the enzyme reaction against β-glucan (Zurbriggen et al., 1990a) and hydroxyethyl cellulose, HEC (IUPAC, 1987). The β-glucanase activity was 0.7 nkat ml^{-1} and

endoglucanase (HEC) activity less than 0.4 nkat ml^{-1}. No endoglucanase activity could be detected in culture filtrates of control cultivations of yeast missing the EG V gene (*S. cerevisiae* DBY 746 carrying the plasmid pAJ 401).

The purified EGV preparation can be obtained from the ultrafiltration concentrate by standard protein chromatography methods. The EGV protein can be bound to an anion exchanger resin (e.g. Mono Q columns or DEAE Sepharose FF, Pharmacia) in low ionic strength buffer and at appropriate pH. The protein can be eluted out of the column using increasing gradient of NaCl (e.g. from 0 to 0.5 M in the buffer of binding). Alternatively, the impurities from the preparation of EGV can be removed by binding them in anion exchange resin at appropriate pH and ionic strength where EGV is not bound to the resin. Cation exchanger resins (e.g. Mono S columns or CM Sepharose FF, Pharmacia) can be used in analogous way by selecting buffers of appropriately low pH (e.g. pH 4 - pH 6). EGV can also be purified by gel permeation chromatography where it can be separated due to its small molecular size. The columns of various materials (e.g. Sephacryl S-100 HR or various types of Sepharose and Superose, Pharmacia; Fractogel TSK HW-55, Merck) in e.g. phosphate or acetate buffers containing e.g. 0.05 - 0.5 M NaCl can be used. Hydrophobic interaction chromatography and various affinity chromatography methods may also be used.

Example 7

Enzyme preparation containing EG V protein obtained from *Trichoderma reesei* grown on glucose.

One of the best producing *T. reesei* QM9414 transformants (number QM/101c) was grown in a bioreactor (Chemap LF 20, working volume 16 l) on a medium of Mandels and Weber (1969) where Solka floc cellulose (10 g l^{-1}) was replaced by 20 g l^{-1} of glucose and where the concentrations of other nutrients were correspondively doubled. The inoculum (5 times 200 ml) was grown in shake flasks in a medium containing 40 g l^{-1} glucose and the adequate mineral salts for nutrients and buffering of the medium. Cultivation conditions were: temperature 29 °C, pH controlled between 4.0 and 5.0, aeration about 15 l min^{-1} and cultivation time 93 h. During the cultivation glucose concentration in the fermentor was maintained above 5 g

l^{-1} by adding continuously sterile glucose (40 g l^{-1}) solution. The mycelium was separated from the medium by centrifugation and the culture supernatant was concentrated 1.6 times by ultrafiltration (PCI ES 625 membranes).

The clarified supernatant was first fractionated by hydrophobic interaction chromatography. The pH of the sample was was adjusted to pH 6.0 and conductivity of the sample to the value corresponding to 10 mM sodium phosphate buffer, pH 6.0, containing 1.25 mol l^{-1} $(NH_4)_2SO_4$. The sample was applied to a column (113 x 110 mm) of Phenyl Sepharose FF (Pharmacia), previously equilibrated with 10 mM sodium phosphate buffer, pH 6.0, containing 1.25 mol l^{-1} $(NH_4)_2SO_4$. Elution was started by the equilibrating buffer followed by a linear decreasing gradient of ammonium sulphate from 1.25 M to 0 M. Fractions (each 450 ml) which contained the major endoglucanase activity were combined, eluted at the end of the decreasing gradient and by 10 mM phosphate buffer. The other adsorbed proteins were eluted by distilled water and the column was washed with 6 M urea.

The enzyme preparation obtained in the first chromatographic step was equilibrated to 4 mM sodium phosphate, pH 7.2 by gel filtration (Sephadex G-25 coarse). The equilibrated protein solution was applied to a column (113 x 190 mm) of DEAE Sepharose FF (Pharmacia), pre-equilibrated with the same buffer. Elution was performed first with the equilibrating buffer to remove unadsorbed proteins and thereafter by stepwise additions of sodium chloride to concentration of 200 mM. Fractions (each 900 ml) which contained the endoglucanase and which eluted by 200 mM NaCl were collected and the fraction with the highest activity was concentrated by ultrafiltration (Amicon PM-10 membranes). The specific activity of the preparate was 360 nkat mg^{-1} protein and purification factor of ca. 30 was obtained when compared to the supernatant of the fermentor liquid.

The preparate was further characterized by isoelectric focusing on PBE-94 anion exchange material (Pharmacia). The column was equilibrated by 25 mM imidazole-HCl buffer, pH 7.4 and elution was carried out by Polybuffer 74 (Pharmacia) - HCl buffer, pH 4.0 according to the manufacturer's instructions. EGV, measured by β-glucanase activity, eluted from the column at pH 6.6 - 7.2.

Example 8

pH-Optimum of EGV

1.0 % solutions of barley β-glucan (Megazyme, Australia) was prepared in McIlvaine buffers, diluted one to four (corresponding ca. 50 mM citrate-phosphate buffer), in pH range from 3.0 to 8.0. Activity of an enzyme sample prepared as described in Example 7 was assayed using as substrate the β-glucan solutions prepared in the varying pH-values. The assay procedure was otherwise similar to the procedure of endoglucanase assay (IUPAC, 1987). Incubation time in the assay was 10 min at 50 °C, after which the enzyme reaction was terminated by boiling. Reducing sugar groups formed in the reaction were measured by DNS-reaction.

The pH-optimum of EGV was 6.0 - 6.5 (Figure 8).

Example 9

pH-Stability of EGV

An EGV sample was prepared as described in Example 7, except that the last concentration by ultrafiltration was omitted (activity 48 nkat/ml, assayed at pH 6.3 against barley β-glucan, analogously to endoglucanase assay, IUPAC, 1987). This sample was diluted (1 part per 2 parts of buffer) by 100 mM buffers of sodium acetate and sodium phosphate, prepared in different pH values. The diluted samples were incubated at 40 °C for 20 h, and the activity was assayed as described earlier. The pH of incubation was measured after the incubation.

More than 80 % of the original activity was observed in the samples incubated at pH range from ca. pH 5.4 to ca. pH 6.8. The relative recovered activity is presented in Figure 9.

Example 10

Hydrolysis of insoluble cellulosic substrates by EGV.

Avicel (Serva 14204) which is mainly crystalline cellulose and phosphoric acid-swollen

amorphous (Walseth, 1952) cellulose were hydrolysed by the new endoglucanase enzyme (EGV) alone and in combinations with the previously known cellobiohydrolases of *T. reesei*. The preparation of EGV was obtained as described in Example 7. CBHI and CBHII were purified from culture filtrates of *Trichoderma reesei* and were pure proteins as judged by SDS-PAGE. As a control, the cellobiohydrolases were incubated without the addition of EGV. The reducing sugars liberated in the treatments were assayed using the DNS method and the reaction products were analysed by HPLC.

The substrates for the hydrolysis were prepared in 50 mM sodium citrate buffer, pH 5.8 in concentration of 10 g l^{-1}. EGV was dosed on the basis of activity against β-glucan at pH 5.8 (500 or 2000 nkat g^{-1} substrate) and cellobiohydrolases (CBHI and CBHII) on the basis of protein (1.0 or 4 mg g^{-1} substrate). The reaction mixtures were incubated for 20 h at pH 5.8 at 40 °C after which the hydrolysis was terminated by boiling. The values for reducing sugars as glucose assayed from the reaction mixture are presented in Tables 3 and 4. The enzyme dosage was 500 nkat g^{-1} substrate for EGV and 1.0 g g^{-1} substrate for CBHI and CBHII (Table 3), and 2000 nkat g^{-1} substrate for EGV and 4.0 g g^{-1} substrate for CBHI and CBHII (Table 4). Duration of the hydrolysis was 20 h in both cases.

The major hydrolysis product of EGV was cellobiose but also cellotetraose was detected in the hydrolysate by HPLC. The strong synergy of EGV with CBHI in the hydrolysis of these substrates can be clearly seen. Even though EGV released only small amounts of soluble sugars the enhancing effect on the cellulose hydrolysis by CBHI was remarkable.

Table 3. Reducing sugars liberated by cellulases from *T. reesei* in the hydrolysis of crystalline cellulose (Avicel) and amorphous (Walseth) cellulose

enzymes	reducing sugars as glucose (mg ml^{-1})	
	Avicel	Walseth
EGV alone	0.00	0.01 (*)
CBHI alone	0.09	0.16

CBHII alone	0.23	0.67
CBHI and EGV	0.15	0.25
CBHII and EGV	0.23	0.74

(*) determined by HPLC

Table 4. Reducing sugars liberated by cellulases from *T. reesei* in the hydrolysis of crystalline cellulose (Avicel) and amorphous (Walseth) cellulose

enzymes	reducing sugars as glucose (mg ml^{-1})	
	Avicel	Walseth
EGV alone	0.03 (*)	0.05
CBHI alone	0.35	0.49
CBHII alone	0.43	1.44
CBHI and EGV	0.43	0.71
CBHII and EGV	0.47	1.67
EGV (larger dosage (**)	0.06	0.13

(*) determined by HPLC
(**) dosed activity 5000 nkat g^{-1} substrate

Example 11

Hydrolysis of soluble cellulosic substrates by EGV

HEC (hydroxyethyl cellulose, Fluka 54290) which is a soluble substituted cellulose polymer and barley β-glucan (Megazyme, Australia) were hydrolysed by the new endoglucanase enzyme (EGV) alone and in combinations with two previously known endoglucanases of *T. reesei*. The preparation of EGV was obtained as described in Example 7. Endoglucanases EGI and EGII were purified from culture filtrates of *Trichoderma reesei* and were pure proteins as

judged by SDS-PAGE. As a control, the endoglucanases were incubated without the addition of EGV. The reducing sugars liberated in the treatments were assayed using the DNS method.

The substrates for the hydrolysis were prepared in 50 mM sodium acetate buffer, pH 5.8 in concentration of 10 g l^{-1}. Endoglucanases were dosed on the basis of activity against β-glucan (dosage for each: 100 nkat g^{-1} substrate) The reaction mixtures were incubated at 40 °C after which the hydrolysis was terminated by boiling. The values for reducing sugars as glucose assayed from the reaction mixture for HEC are presented in Table 5, and for β-glucan in Table 6. The enhancing effect of EGV on the hydrolysis of HEC and especially β-glucan can clearly be seen.

Table 5. Reducing sugars liberated by endoglucanases from *T. reesei* in the hydrolysis of hydroxyethyl cellulose.

enzymes	hydrolysis products as reducing sugars (mg ml^{-1})	
	2 h hydrolysis	20 h hydrolysis
EGV alone	0.00	0.00
EGI alone	0.14	0.26
EGII alone	0.18	0.29
EGI and EGV	0.16	0.27
EGII and EGV	0.24	0.29

Table 6. Reducing sugars liberated by endoglucanases from *T. reesei* in the hydrolysis of barley β-glucan. Duration of hydrolysis 2 h.

enzymes	hydrolysis products as reducing sugars (mg ml^{-1})
EGV alone	0.96
EGI alone	1.1
EGII alone	0.90
EGI and EGV	2.4
EGII and EGV	2.0

Example 12

Modelling of the cellulose-binding domain of EGV

Sequences of the five cellulose-binding domains (CBD) of the *T. reesei* cellulases were aligned to define conserved regions using the computer program MALIGN (Johnson *et al.*, 1993; Johnson and Overington, 1993), which is suitable for this purpose because the percentage identity between the five CBDs is ~60 %. The construction of a 3-D model of the EGV CBD was performed using the COMPOSER method (Sutcliffe *et al.*, 1987a,b; Blundell *et al.*, 1988; Sali *et al.*, 1990), which is based on rules derived from known three-dimensional structures. These rules can be used to define a conserved core for the model, to select appropriate fragments for the variable regions and to replace the side chains. The NMR-based structure of the CBHI CBD (Kraulis *et al.*, 1989) was used as a basis for the EGV model. The computer program CHARMm ver. 22 (Brooks *et al.*, 1983) was used to soak the completed model in a 35 Å cubic box of water and to refine the model through energy minimization and molecular dynamics simulation of 100 ps under periodic boundary conditions.

The sequence alignment shows that the CBDs of *T. reesei* are highly conserved except for one insertion and one deletion of a single aa in EGV. Therefore most parts of the 3-D structure of the CBHI CBD, determined by NMR (Kraulis *et al.*, 1989), could be used as a conserved core

for modelling of the EGV CBD by the computer program COMPOSER. The CBHI CBD is a wedge-shaped domain having two flat surfaces. One of these is predominantly hydrophilic and contains three tyrosine residues that have been shown by chemical modification to be important for the binding of the enzyme to cellulose (Claeyssens and Tomme, 1989). Tyr^{492} located at the tip of the wedge has also been demonstrated by site-directed mutagenesis to be involved in substrate binding (Reinikainen *et al.*, 1992). This residue is replaced by a tryptophan (Trp^{236}) in the EGV CBD (Fig 3B), an amino acid substitution also seen in many other fungal CBDs. Both tyrosine and tryptophan residues interact readily with carbohydrates.

The backbones of the CBHI and EGV CBDs are very similar. Two disulfide bridges in identical positions stabilize the structures. The insertion and the deletion in EGV are situated in a single loop and thus compensate each other, maintaining the loop backbone unchanged compared with that of CBHI. However, there is an interesting difference in the backbone conformation at the other, more hydrophobic, flat face. A substantial change in torsion angle was observed at position Gly^{220} of EGV, where the ϕ-angle of the glycine residue changes from negative to positive during the refinement simulation. This causes the loop at region 217-221 to be pushed outwards. Interestingly, the corresponding loop in the CBHI CBD contains a proline residue (Pro^{474}), mutation of which reduces the activity of CBHI against crystalline cellulose (Reinikainen *et al.*, 1992). Simulations of the other CBDs from *T. reesei* show that this region is the most flexible region of the CBD (A.-M. Hoffrén, T.T. Teeri, and O. Teleman, submitted).

There are also moderate differences in the backbone regions, in which proline (Pro^{491}) and serine (Ser^{482}) residues of the CBHI CBD are replaced by glutamine (Gln^{235}) and proline (Pro^{227}) residues in the EGV CBD, respectively. Moderate differences are also found in side chain conformations, but these changes are within the range of fluctuation occurring during simulation. One of the tyrosines (Tyr^{210}) forming the hydrophilic face of EGV CBD points more upwards than its counterpart (Tyr^{466}) in CBHI. This difference in orientation is significant, but the flexibility allows Tyr^{210} in EGV to occupy the same position as Tyr^{466} in CBHI. Thus the difference in orientation is unlikely to affect substantially the affinity for cellulose. Overall, the EGV CBD seems to be less wedge-shaped and the hydrophobic surface more rounded than that of the CBHI CBD.

REFERENCES

Ballance, J.D. (1991), Transformation systems for filamentous fungi and an overview of fungal gene structure. In *Molecular Industrial Mycology. Systems and Applications for Filamentous Fungi*. Leon, S.A., and Berka, R.M. (eds.). New York: Marcel Dekker, Inc., pp. 1-29.

Béguin, P., Gilkes, N.R., Kilburn, D.G., Miller, R.C.Jr., O'Neill, G. and Warren, R.A.J. (1987), Cloning of cellulase genes. *CRC Crit. Rev. Biotechnol.* 6: 129-162.

Blundell, T.L., Carney, D., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J.P., Singh, D.A., Sibanda, B.L., and Sutcliffe, M. (1988), Knowledge-based protein modelling and design. *Eur J Biochem* **172**: 513-520.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., Sates, D.J., Swaminathan, S., and Karplus, M. (1983), CHARMm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4: 187-217.

Chen, C.M., Gritzali, M., and Stafford, D.W. (1987), Nucleotide sequence and deduced primary structure of cellobiohydrolase II from *Trichoderma reesei*. **Bio/Technology 5**: 274-278.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979), Isolation of biologically active ribonucleic acid from sources rich in ribonuclease. *Biochem J* 18: 5294-5299.

Claeyssens, M., and Tomme, P. (1989), Structure-function relationships of cellulolytic proteins from *Trichoderma reesei*. In *Trichoderma reesei* Cellulases: Biochemistry, Genetics, Physiology and Application. Kubicek, C.P. *et al.* (eds.). Cambridge: Royal Society of Chemistry, pp. 1-11.

Gong, C.-S., Ladisch, M.R. and Tsao, G.T. (1979), Biosynthesis, purification, and mode of

action of cellulases of *Trichoderma reesei*. *Adv. Chem. Ser.* 181: 261-287.

Gritz, L. and Davies, J. (1983), *Gene* 25: 179 - 188.

Hanahan, H. (1983), *J. Mol. Biol.* 166: 557-580.

Hynes, M.J., Corrick, C.M. and King, J.A. (1983). *Mol. Cell. Biol.* 3: 1430-1439.

Håkansson, U., Fägerstam, L., Pettersson, G., and Andersson, L. (1978) Purification and characterization of a low molecular weight 1,4-ß-glucan glucanohydrolase from the cellulolytic fungus *Trichoderma viride* QM 9414. *Biochim Biophys Acta* 524: 385-392.

IUPAC (International Union of Pure and Applied Chemistry) (1987) *Pure Appl. Chem.* 59: 257-268.

Karhunen et al. (1993) High frequency one-step gene replacement in *Trichoderma reesei* I, Endoglucanase I overproduction, *Mol Gen Genet* 241: 515-522.

Kraulis, P.J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J. and Gronenborn, A.M. (1989) *Biochemistry* 28: 7241-7257.

Mach, R.L., M. Schindler and C.P. Kubicek. (1994), *Curr. Genet.* 25:567-570.

Mandels, M. and Weber, J. (1969), Production of cellulases. *Adv Chem. Ser.* 95: 391-413.

Minet, M. and Lacroute, F. (1990), Cloning and sequencing of a human cDNA coding for a multifunctional polypeptide of the purine pathway by complementation of the ade2-101 mutant in *Saccharomyces cerevisiae*. *Curr. Genet.* 18: 287-291.

Nakari, T.. Alatalo, E. and Penttilä, M. (1992) Poster presentation at ECFG 1, Nottingham, England, 1992

Nevalainen, K.M.H., Penttilä, M.E., Harkki, A., Teeri, T.T. and Knowles J. (1991), *In*: Leong, S.A. and Berka, R.M. (eds.) *Mol. Ind. Mycology*, Marcel Dekker, Inc., New York, pp. 129 - 148.

Ooi, T., Shinmyo, A., Okada, H., Hara, S., Ikenaka, T., Murao, S., and Arai, M. (1990) Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from *Aspergillus aculeatus*. *Curr. Genet.* **18**: 217-222.

Penttilä, M., Lehtovaara, P., Nevalainen, H., Bhikhabhai, R. and Knowles, J. (1986) Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. *Gene* **45**: 253-263.

Penttilä, M.E., André, L., Saloheimo, M., Lehtovaara, P. and Knowles, J.K.C (1987a) Expression of two *Trichoderma reesei* endoglucanases in the yeast *Saccharomyces cerevisiae*. *Yeast* **3**: 175-185.

Penttilä, M., Nevalainen, K.M.H., Rättö, M., Salminen, E. and Knowles, J.K.C. (1987b), *Gene* **61**: 155-164.

Penttilä, M.E., André, L., Lehtovaara, P., Bailey, M., Teeri, T.T. and Knowles, J.K.C. (1988) Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. *Gene* **63**: 103-112.

Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T.A., Knowles, J.K.C. and Teeri, T.T. (1992), Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. *PROTEINS* **14**: 475-482.

Sali, A., Overington, J.P., Johnson, M.S., and Blundell, T.L. (1990), From comparisons of protein sequences and structures to protein modelling and design. *TIBS* **15**: 235-240.

Saloheimo, M., Lehtovaara, P., Penttilä, M., Teeri, T.T., Ståhlberg, J., Johansson, G., Pettersson, G., Claeyssens, M.. Tomme, P., and Knowles, J.K.C. (1988), EGIII, a new

endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. *Gene* **63**: 11-21.

Schindler, M., R.L. Mach, S.K. Vollerhofer, R. Hodits, R. Guber, J. Visser, L. De Graaff and C.P. Kubicek (1993). *Gene* **130**: 271-275.

Sherman, F. (1991) Getting started with yeast. In *Guide to Yeast Genetics and Molecular Biology*. Guthrie, C., and Fink, G.R. (eds.). *Methods in Enzymology*, Vol 194. California: Academic Press Inc., pp. 3-21.

Shoemaker, S., Schweickart, V., Ladner, M., Gelfand, D., Kwok, S., Myambo, K., and Innis, M. (1983) Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. *Bio/Technology* **1**: 691-695.

Ståhlberg, J. (1991) Functional organization of cellulases from *Trichoderma reesei*. PhD Thesis, Inst. of Biochemistry, University of Uppsala, Sweden, 1991.

Suominen et al., (1993, High frequency one-step gene replacement in *Trichoderma* reesei II, Effects of deletions of individual cellulase genes. *Mol Gen Genet* **241**: 523-530.

Sutcliffe, J.G. (1979), Cold Spring Harbor Symp. Wuant. Biol. **43**: 77-90

Sutcliffe, M.J., Hanef, I., Carney, D., and Blundell, T.L. (1987a), Knowledge based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Engin* **1**: 377-384.

Sutcliffe, M.J., Hayes, F.R.F., and Blundell, T.L. (1987b), Knowledge based modelling of homologous proteins, part II: rules for the conformations of substituted side chains. *Protein Engin* **1**: 385-392.

Teeri, T., Salovuori, I., and Knowles, J. (1983), The molecular cloning of the major cellulase gene from *Trichoderma reesei*. *Bio/Technology* **1**: 696-699.

Teeri, T.T., Lehtovaara, P., Kauppinen, S., Salovuori, I., and Knowles, J. (1987) Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. *Gene* **51**: 43-52.

Ülker, A. and Sprey, B. (1990), Production and characterization of an unglycosylated low molecular weight 1,4-ß-glucan-glucanohydrolase of *Trichoderma reesei*. In Kubicek, C.P. and Kubicek-Pranz, E.M. (eds.), *Trichoderma Cellulases: Biochemistry, Physiology, Genetics and Applications*. Royal Chemical Society Press, Cambridge, 1990, pp. 60-79.

van Arsdell, J.N., Kwok, S., Schweickart, V.L., Ladner, M.B., Gelfand, D.H. and Innis, M.A. (1987), Cloning, characterization, and expression in *Saccharomyces cerevisiae* of endoglucanase I from *Trichoderma reesei*. Bio/Technology **5**: 60-64.

van Gorcom et al. (1985), *Gene* **40**: 99-106

Walseth, C.S. (1952) *Tappi* **35**: 228-233.

Zurbriggen, B., Bailey, M., Penttilä, M., Poutanen, K. and Linko, M. (1990a) *J. Biotechnol.* **17**: 133-146.

Zurbriggen, B., Bailey, M., Penttilä, M., Poutanen, K. and Linko, M. (1990b), *J. Biotechnol.* **13**: 267-278.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: OY ALKO AB
(B) STREET: Salmisaarenranta 7 H
(C) CITY: Helsinki
(E) COUNTRY: Finland
(F) POSTAL CODE: FIN-00180

(ii) TITLE OF INVENTION: Novel Endoglucanase Enzyme

(iii) NUMBER OF SEQUENCES: 10

(iv) COMPUTER READABLE FORM
(A) MEDIUM TYPE: Floppy disc
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Releae #1.0, Version #1.25 (EPO)

(v) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: FI 932521
(B) FILING DATE: 2-JUNE-1993

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 884 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Trichoderma reesei
(B) STRAIN: QM9414

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 40..765

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GATCTTCCAT CTCGTGTCTT GCTTGTAACC ATCGTGACC ATG AAG GCA ACT CTG          54
                                            Met Lys Ala Thr Leu
                                              1               5

GTT CTC GGC TCC CTC ATT GTA GGC GCC GTT TCC GCG TAC AAG GCC ACC        102
Val Leu Gly Ser Leu Ile Val Gly Ala Val Ser Ala Tyr Lys Ala Thr
                 10                  15                  20

ACC ACG CGC TAC TAC GAT GGG CAG GAG GGT GCT TGC GGA TGC GGC TCG        150
Thr Thr Arg Tyr Tyr Asp Gly Gln Glu Gly Ala Cys Gly Cys Gly Ser
             25                  30                  35

AGC TCC GGC GCA TTC CCG TGG CAG CTC GGC ATC GGC AAC GGA GTC TAC        198
Ser Ser Gly Ala Phe Pro Trp Gln Leu Gly Ile Gly Asn Gly Val Tyr
         40                  45                  50
```

```
ACG GCT GCC GGC TCC CAG GCT CTC TTC GAC ACG GCC GGA GCT TCA TGG      246
Thr Ala Ala Gly Ser Gln Ala Leu Phe Asp Thr Ala Gly Ala Ser Trp
     55                  60                  65

TGC GGC GCC GGC TGC GGT AAA TGC TAC CAG CTC ACC TCG ACG GGC CAG      294
Cys Gly Ala Gly Cys Gly Lys Cys Tyr Gln Leu Thr Ser Thr Gly Gln
 70                  75                  80                  85

GCG CCC TGC TCC AGC TGC GGC ACG GGC GGT GCT GCT GGC CAG AGC ATC      342
Ala Pro Cys Ser Ser Cys Gly Thr Gly Gly Ala Ala Gly Gln Ser Ile
             90                  95                 100

ATC GTC ATG GTG ACC AAC CTG TGC CCG AAC AAT GGG AAC GCG CAG TGG      390
Ile Val Met Val Thr Asn Leu Cys Pro Asn Asn Gly Asn Ala Gln Trp
        105                 110                 115

TGC CCG GTG GTC GGC GGC ACC AAC CAA TAC GGC TAC AGC TAC CAT TTC      438
Cys Pro Val Val Gly Gly Thr Asn Gln Tyr Gly Tyr Ser Tyr His Phe
    120                 125                 130

GAC ATC ATG GCG CAG AAC GAG ATC TTT GGA GAC AAT GTC GTC GTC GAC      486
Asp Ile Met Ala Gln Asn Glu Ile Phe Gly Asp Asn Val Val Val Asp
135                 140                 145

TTT GAG CCC ATT GCT TGC CCC GGG CAG GCT GCC TCT GAC TGG GGG ACG      534
Phe Glu Pro Ile Ala Cys Pro Gly Gln Ala Ala Ser Asp Trp Gly Thr
150                 155                 160                 165

TGC CTC TGC GTG GGA CAG CAA GAG ACG GAT CCC ACG CCC GTC CTC GGC      582
Cys Leu Cys Val Gly Gln Gln Glu Thr Asp Pro Thr Pro Val Leu Gly
                170                 175                 180

AAC GAC ACG GGC TCA ACT CCT CCC GGG AGC TCG CCG CCA GCG ACA TCG      630
Asn Asp Thr Gly Ser Thr Pro Pro Gly Ser Ser Pro Pro Ala Thr Ser
        185                 190                 195

TCG AGT CCG CCG TCT GGC GGC GGC CAG CAG ACG CTC TAT GGC CAG TGT      678
Ser Ser Pro Pro Ser Gly Gly Gly Gln Gln Thr Leu Tyr Gly Gln Cys
    200                 205                 210

GGA GGT GCC GGC TGG ACG GGA CCT ACG ACG TGC CAG GCC CCA GGG ACC      726
Gly Gly Ala Gly Trp Thr Gly Pro Thr Thr Cys Gln Ala Pro Gly Thr
215                 220                 225

TGC AAG GTT CAG AAC CAG TGG TAC TCC CAG TGT CTT CCT TGAGAAGGCC       775
Cys Lys Val Gln Asn Gln Trp Tyr Ser Gln Cys Leu Pro
230                 235                 240

CAAGATAGCC ATGTCTCTCT AGCATTCTTC CGGCGTCAGT CTGATCTGCC TATTTAATCA    835

GGTCAGTCAA TATGTATCCA GAGATAATAA ATTATGTATA TTATAGCAG                884
```

(2) INFORMATION FOR SEQ ID NO: 2

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 166 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Lys Ala Thr Leu Val Leu Gly Ser Leu Ile Val Gly Ala Val Ser
 1               5                  10                  15
```

```
Ala Tyr Lys Ala Thr Thr Thr Arg Tyr Tyr Asp Gly Gln Glu Gly Ala
                20                  25                  30

Cys Gly Cys Gly Ser Ser Ser Gly Ala Phe Pro Trp Gln Leu Gly Ile
            35                  40                  45

Gly Asn Gly Val Tyr Thr Ala Ala Gly Ser Gln Ala Leu Phe Asp Thr
        50                  55                  60

Ala Gly Ala Ser Trp Cys Gly Ala Gly Cys Gly Lys Cys Tyr Gln Leu
65                  70                  75                  80

Thr Ser Thr Gly Gln Ala Pro Cys Ser Ser Cys Gly Thr Gly Gly Ala
                85                  90                  95

Ala Gly Gln Ser Ile Ile Val Met Val Thr Asn Leu Cys Pro Asn Asn
            100                 105                 110

Gly Asn Ala Gln Trp Cys Pro Val Val Gly Gly Thr Asn Gln Tyr Gly
        115                 120                 125

Tyr Ser Tyr His Phe Asp Ile Met Ala Gln Asn Glu Ile Phe Gly Asp
    130                 135                 140

Asn Val Val Val Asp Phe Glu Pro Ile Ala Cys Pro Gly Gln Ala Ala
145                 150                 155                 160

Ser Asp Trp Gly Thr Cys Leu Cys Val Gly Gln Gln Glu Thr Asp Pro
                165                 170                 175

Thr Pro Val Leu Gly Asn Asp Thr Gly Ser Thr Pro Pro Gly Ser Ser
            180                 185                 190

Pro Pro Ala Thr Ser Ser Ser Pro Pro Ser Gly Gly Gly Gln Gln Thr
        195                 200                 205

Leu Tyr Gly Gln Cys Gly Gly Ala Gly Trp Thr Gly Pro Thr Thr Cys
    210                 215                 220

Gln Ala Pro Gly Thr Cys Lys Val Gln Asn Gln Trp Tyr Ser Gln Cys
225                 230                 235                 240

Leu Pro
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 166 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
Tyr Lys Ala Thr Thr Thr Arg Tyr Tyr Asp Gly Gln Glu Gly Ala Cys
1               5                  10                  15

Gly Cys Gly Ser Ser Ser Gly Ala Phe Pro Trp Gln Leu Gly Ile Gly
            20                  25                  30
```

Asn Gly Val Tyr Thr Ala Ala Gly Ser Gln Ala Leu Phe Asp Thr Ala
35 40 45

Gly Ala Ser Trp Cys Gly Ala Gly Cys Gly Lys Cys Tyr Gln Leu Thr
50 55 60

Ser Thr Gly Gln Ala Pro Cys Ser Ser Cys Gly Thr Gly Gly Ala Ala
65 70 75 80

Gly Gln Ser Ile Ile Val Met Val Thr Asn Leu Cys Pro Asn Asn Gly
85 90 95

Asn Ala Gln Trp Cys Pro Val Val Gly Gly Thr Asn Gln Tyr Gly Tyr
100 105 110

Ser Tyr His Phe Asp Ile Met Ala Gln Asn Glu Ile Phe Gly Asp Asn
115 120 125

Val Val Val Asp Phe Glu Pro Ile Ala Cys Pro Gly Gln Ala Ala Ser
130 135 140

Asp Trp Gly Thr Cys Leu Cys Val Gly Gln Gln Glu Thr Asp Pro Thr
145 150 155 160

Pro Val Leu Gly Asn Asp
165

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2211 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GAATTCTCAC GGTGAATGTA GGCCTTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA 60

ACCTCCATTA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCATGGCAC TGTTCTCAAA 120

TAGATTGGGG AGAAGTTGAC TTCCGCCCAG AGCTGAAGGT CGCACAACCG CATGATATAG 180

GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTTG CGATCTAACA 240

TCCAGGAACC TGGATACATC CATCATCACG CACGACCACT TTGATCTGCT GGTAAACTCG 300

TATTCGCCCT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGCCCCTTTC GGTATACTGC 360

GTGTGTCTTC TCTAGGTGCA TTCTTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGGAG 420

TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC 480

TGCATCATGT ATATAATAGT GATCCTGAGA AGGGGGGTTT GGAGCAATGT GGGACTTTGA 540

TGGTCATCAA ACAAAGAACG AAGACGCCTC TTTTGCAAAG TTTTGTTTCG GCTACGGTGA 600

AGAACTGGAT ACTTGTTGTG TCTTCTGTGT ATTTTTGTGG CAACAAGAGG CCAGAGACAA 660

TCTATTCAAA CACCAAGCTT GCTCTTTTGA GCTACAAGAA CCTGTGGGGT ATATATCTAG 720

AGTTGTGAAG TCGGTAATCC CGCTGTATAG TAATACGAGT CGCATCTAAA TACTCCGAAG 780

CTGCTGCGAA CCCGGAGAAT CGAGATGTGC TGGAAAGCTT CTAGCGAGCG GCTAAATTAG 840

CATGAAAGGC TATGAGAAAT TCTGGAGACG GCTTGTTGAA TCATGGCGTT CCATTCTTCG 900

ACAAGCAAAG CGTTCCGTCG CAGTAGCAGG CACTCATTCC CGAAAAAACT CGGAGATTCC 960

TAAGTAGCGA TGGAACCGGA ATAATATAAT AGGCAATACA TTGAGTTGCC TCGACGGTTG 1020

CAATGCAGGG GTACTGAGCT TGGACATAAC TGTTCCGTAC CCCACCTCTT CTCAACCTTT 1080

GGCGTTTCCC TGATTCAGCG TACCCGTACA AGTCGTAATC ACTATTAACC CAGACTGACC 1140

GGACGTGTTT TGCCCTTCAT TTGGAGAAAT AATGTCATTG CGATGTGTAA TTTGCCTGCT 1200

TGACCGACTG GGGCTGTTCG AAGCCCGAAT GTAGGATTGT TATCCGAACT CTGCTCGTAG 1260

AGGCATGTTG TGAATCTGTG TCGGGCAGGA CACGCCTCGA AGGTTCACGG CAAGGGAAAC 1320

CACCGATAGC AGTGTCTAGT AGCAACCTGT AAAGCCGCAA TGCAGCATCA CTGGAAAATA 1380

CAAACCAATG GCTAAAAGTA CATAAGTTAA TGCCTAAAGA AGTCATATAC CAGCGGCTAA 1440

TAATTGTACA ATCAAGTGGC TAAACGTACC GTAATTTGCC AACGCGTTGT GGGGTTGCAG 1500

AAGCAACGGC AAAGCCCACT TCCCACGTTT GTTTCTTCAC TCAGTCCAAT CTCAGCTGGT 1560

GATCCCCCAA TTGGGTCGCT TGTTTGTTCC GGTGAAGTGA AAGAAGACAG AGGTAAGAAT 1620

GTCTGACTCG GAGCGTTTTG CATACAACCA AGGGCAGTGA TGGAAGACAG TGAAATGTTG 1680

ACATTCAAGG AGTATTTAGC CAAGGGATGCTTGAGTGTATC GTGTAAGGAG GTTTGTCTGC 1740

CGATACGACG AATACTGTAT AGTCACTTCT GATGAAGTGG TCCATATTGA AATGTAAGTC 1800

GGCACTGAAC AGGCAAAAGA TTGAGTTGAA ACTGCCTAAG ATCTCGGGCC CTCGGGCTTC 1860

GGCTTTGGGT GTACATGTTT GTGCTCCGGG CAAATGCAAA GTGTGGTAGG ATCGACACAC 1920

TGCTGCCTTT ACCAAGCAGC TGAGGGTATG TGATAGGCAA ATGTTCAGGG GCCACTGCAT 1980

GGTTTCGAAT AGAAAGAGAA GCTTAGCCAA GAACAATAGC CGATAAAGAT AGCCTCATTA 2040

AACGAAATGA GCTAGTAGGC AAAGTCAGCG AATGTGTATA TATAAAGGTT CGAGGTCCGT 2100

GCCTCCCTCA TGCTCTCCCC ATCTACTCAT CAACTCAGAT CCTCCAGGAG ACTTGTACAC 2160

CATCTTTTGA GGCACAGAAA CCCAATAGTC AACCGCGGAC TGCGCATCAT G 2211

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1627 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGCGGTATTG GCTACAGCGG CCCCACGGTC TGCGCCAGCG GCACAACTTG CCAGGTCCTG 60

AACCCTTACT ACTCTCAGTG CCTGTAAAGC TCCGTGCGAA AGCCTGACGC ACCGGTAGAT 120

TCTTGGTGAG CCCGTATCAT GACGGCGGCG GGAGCTACAT GGCCCCGGGT GATTTATTTT 180

TTTTGTATCT ACTTCTGACC CTTTTCAAAT ATACGGTCAA CTCATCTTTC ACTGGAGATG 240

CGGCCTGCTT GGTATTGCGA TGTTGTCAGC TTGGCAAATT GTGGCTTTCG AAAACACAAA 300

ACGATTCCTT AGTAGCCATG CATCGGGATC CTTTAAGATA ACGGAATAGA AGAAAGAGGA 360

AATTAAAAAA AAAAA AAAA CAAACATCCC GTTCATAACC CGTAGAATCG CCGCTCTTCG 420

TGTATCCCAG TACCACGGCA AAGGTATTTC ATGATCGTTC AATGTTGATA TTGTTCCCGC 480

CAGTATGGCT GCACCCCCAT CTCCGCGAAT CTCCTCTTCT CGAACGCGGT AGTGGCGCGC 540

CAATTGGTAA TGACCATAGG GAGACAAACA GCATAATAGC AACAGTGGAA ATTAGTGGCG 600

CAATAATTGA GAACACAGTG AGACCATAGC TGGCGGCCTG GAAAGCACTG TTGGAGACCA 660

ACTTGTCCGT TGCGAGGCCA ACTTGCATTG CTGTCAAGAC GATGACAACG TAGCCGAGGA 720

CCGTCACAAG GGACGCAAAG TTGTCGCGGA TGAGGTCTCC GTAGATGGCA TAGCCGGCAA 780

TCCGAGAGTA GCCTCTCAAC AGGTGGCCTT TTCGAAACCG GTAAACCTTG TTCAGACGTC 840

CTAGCCGCAG CTCACCGTAC CAGTATCGAG GATTGACGGC AGAATAGCAG TGGCTCTCCA 900

GGATTTGACT GGACAAAATC TTCCAGTATT CCCAGGTcAC AGTGTCTGGC AGAAGTCCCT 960

TCTCGCGTGC ANTCGAAAGT CGCTATAGTG CGCAATGAGA GCACAGTAGG AGAATAGGAA 1020

CCCGCGAGCA CATTGTTCAA TCTCCACATG AATTGGATGA CTGCTGGGCA GAATGTGCTG 1080

CCTCCAAAAT CCTGCGTCCA ACAGATACTC TGGCAGGGGC TTCAGATGAA TGCCTCTGGG 1140

CCCCCAGATA AGATGCAGCT CTGGATTCTC GGTTACNATG ATATCGCGAG AGAGCACGAG 1200

TTGGTGATGG AGGGACAGGA GGCATAGGTC GCGCAGGCCC ATAACCAGTC TTGCACAGCA 1260

TTGATCTTAC CTCACGAGGA GCTCCTGATG CAGAAACTCC TCCATGTTGC TGATTGGGTT 1320

GAGAATTTCA TCGCTCCTGG ATCGTATGGT TGCTGGCAAG ACCCTGCTTA ACCGTGCCGT 1380

GTCATGGTCA TCTCTGGTGG CTTCGTCGCT GGCCTGTCTT TGCAATTCGA CAGCAAATGG 1440

TGGAGATCTC TCTATCGTGA CAGTCATGGT AGCGATAGCT AGGTGTCGTT GCACGCACAT 1500

AGGCCGAAAT GCGAAGTGGA AAGAATTTCC CGGNTGCGGA ATGAAGTCTC GTCATTTTGT 1560

ACTCGTACTC GACACCTCCA CCGAAGTGTT AATAATGGAT CCACGATGCC AAAAAGCTTG 1620

TGCATGC 1627

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1137 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GAATTCTCAC GGTGAATGTA GGCCTTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA 60

ACCTCCATTA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCATGGCAC TGTTCTCAAA 120

TAGATTGGGG AGAAGTTGAC TTCCGCCCAG AGCTGAAGGT CGCACAACCG CATGATATAG 180

GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTTG CGATCTAACA 240

TCCAGGAACC TGGATACATC CATCATCACG CACGACCACT TTGATCTGCT GGTAAACTCG 300

TATTCGCCCT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGCCCCTTTC GGTATACTGC 360

GTGTGTCTTC TCTAGGTGCA TTCTTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGGAG 420

TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC 480

TGCATCATGT ATATAATAGT GATCCTGAGA AGGGGGGTTT GGAGCAATGT GGGACTTTGA 540

TGGTCATCAA ACAAAGAACG AAGACGCCTC TTTTGCAAAG TTTTGTTTCG GCTACGGTGA 600

AGAACTGGAT ACTTGTTGTG TCTTCTGTGT ATTTTTGTGG CAACAAGAGG CCAGAGACAA 660

TCTATTCAAA CACCAAGCTT GCTCTTTTGA GCTACAAGAA CCTGTGGGGT ATATATCTAG 720

TGGCCAGAAT GCCTAGGTCA CCTCTAGAGA GTTGAAACTG CCTAAGATCT CGGGCCCTCG 780

GGCTTCGGCT TTGGGTGTAC ATGTTTGTGC TCCGGGCAAA TGCAAAGTGT GGTAGGATCG 840

ACACACTGCT GCCTTTACCA AGCAGCTGAG GGTATGTGAT AGGCAAATGT TCAGGGGCCA 900

CTGCATGGTT TCGAATAGAA AGAGAAGCTT AGCCAAGAAC AATAGCCGAT AAAGATAGCC 960

TCATTAAACG AAATGAGCTA GTAGGCAAAG TCAGCGAATG TGTATATATA AAGGTTCGAG 1020

GTCCGTGCCT CCCTCATGCT CTCCCCATCT ACTCATCAAC TCAGATCCTC CAGGAGACTT 1080

GTACACCATC TTTTGAGGCA CAGAAACCCA ATAGTCAACC GCGGACTGCG CATCATG 1137

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGTCTGAAGG ACGTGGAATG ATGG 24

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GATGCATCGA TCGTCCGCGG GTTGAGAGAA GTTGTTGGAT TGATCAAAAA G 51

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAGAGACCGC GGTGATCTTC CATCTCGTGT CTTGCTTGTA AC 42

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

ATCGTGGATC CATTATTAAC ACTTCGGTGG

Claims:

1. An isolated DNA sequence, c h a r a c t e r i z e d in that it codes for a *Trichoderma* enzyme having endoglucanase activity, the molecular weight of the unglycosylated form of said enzyme being 20 to 25 kDa and said enzyme comprising a core domain, a linker region and a cellulose binding domain, and functional parts thereof.

2. The DNA sequence according to claim 1, wherein the DNA sequence hybridizes to the DNA sequence of SEQ ID NO. 1 or to the DNA sequence of SEQ ID NO. 11.

3. The DNA sequence according to claim 1, wherein the DNA sequence codes for the amino acid sequence of SEQ ID NO. 2.

4. The DNA sequence according to claim 1, wherein the DNA sequence is the DNA sequence of SEQ ID NO. 1 or to the DNA sequence of SEQ ID NO. 11.

5. A DNA sequence, which codes for a polypeptide having endoglucanase activity, said sequence coding for the sequence of SEQ ID NO. 3 or functional equivalents thereof.

6. A vector construction, c h a r a c t e r i z e d in that it comprises the DNA sequence of any one of claims 1 to 5.

7. A microorganism host, c h a r a c t e r i z e d in that it has been transformed with the DNA sequence of any one of claims 1 to 5 or with a vector construction of claim 6 and is able to express said DNA sequence.

8. The host according to claim 7, wherein said host is a fungal or yeast host.

9. The host according to claims 7 or 8, wherein said host is *Trichoderma*.

10. The host according to claims 7 or 8, wherein said host is *Saccharomyces*.

11. A culture medium, characterized in that is comprises the enzymes secreted from the host of claims 7 to 10.

12. A product derived from the culture medium of claim 11 by purifying, concentrating, drying or immobilizing said culture medium.

13. A method for isolating a DNA sequence coding for *Trichoderma* enzyme having endoglucanase activity,
characterized by

- enriching the mRNA pool of a *Trichoderma* strain producing endoglucanase activity in respect of the mRNA of the endoglucanase by culturing the *Trichoderma* strain in conditions which will induce the endoglucanase production of said strain,
- isolating mRNA from the strain,
- preparing cDNA corresponding to the isolated mRNA,
- placing the cDNA thus obtained in a vector under the control of a yeast promoter,
- transforming the recombinant plasmids into a yeast strain which naturally does not produce the endoglucanase in order to provide an expression library,
- cultivating the yeast clones thus obtained on a culture medium in order to express the expression library in the yeast,
- separating the yeast clones producing the endoglucanase from the other yeast clones,
- isolating the plasmid-DNA of said separated yeast clones, and,
- if desired, sequencing the DNA in order to determine the DNA sequence coding for the endoglucanase.

14. The method according to claim 13, wherein the recombinant plasmids are transformed into a strain of the yeast *Saccharomyces cerevisiae*.

15. The method according to claim 13, wherein the yeast clones are cultivated on a culture medium containing at least one substrate selected from the group comprising β-glucan, hydroxyethyl cellulose, methylumbelliferyl lactoside and methylumbelliferyl cellobioside.

16. A method for constructing a *Trichoderma* host capable of expressing an endoglucanase

enzyme, c h a r a c t e r i z e d in that it comprises

a) isolating the DNA sequence coding for an endoglucanase, the molecular weight of which in unglycosylated form is 20 to 25 kDa, or parts thereof, from a suitable donor strain,

b) constructing a vector carrying said DNA sequence and

c) transforming the vector obtained into a *Trichoderma* host.

17. A method for producing a *Trichoderma* endoglucanase enzyme, c h a r a c t e r i z e d in that it comprises the steps of

a) isolating the DNA sequence coding for an endoglucanase, the molecular weight of which in unglycosylated form is 20 to 25 kDa, or functional parts thereof, from a suitable donor strain

b) constructing a vector carrying said DNA sequence,

c) transforming the vector obtained to a *Trichoderma* host to obtain a recombinant host strain,

d) cultivating said recombinant host strain under conditions permitting expression of said endoglucanase, and

e) recovering said endoglucanase.

18. A method for constructing a *Saccharomyces* host capable of expressing an endoglucanase enzyme, c h a r a c t e r i z e d in that it comprises

a) isolating the DNA sequence coding for an endoglucanase, the molecular weight of whose unglycosylated form being 20 to 25 kDa, and functional parts thereof, from a suitable donor strain,

b) constructing a vector carrying said DNA sequence and

c) transforming the vector obtained into a *Saccharomyces* host.

19. A method for producing a *Trichoderma* endoglucanase enzyme, c h a r a c t e r i z e d by

a) isolating the DNA sequence coding for an endoglucanase, the molecular weight of which is 20 to 25 kDa in unglycosylated form, or functional parts thereof, from a suitable donor strain

b) constructing a vector carrying said DNA sequence,

c) transforming the vector obtained to a *Saccharomyces* host to obtain a recombinant host strain,

d) cultivating said recombinant host strain under conditions permitting expression of said endoglucanase, and

e) recovering said endoglucanase.

20. An enzyme preparation, c h a r a c t e r i z e d in that it contains an endoglucanase enzyme having the amino acid sequence of SEQ ID NO. 2 or functional derivatives thereof.

21. An enzyme preparation, c h a r a c t e r i z e d in that it contains elevated levels of an endoglucanase enzyme having in unglycosylated form a molecular weight in the range from 20 to 25 kDa, or functional parts thereof, and exhibiting catalytic activity towards the substrates β-glucan.

22. The enzyme preparation according to claim 22, wherein the endoglucanase enzymes exhibits activity towards crystalline cellulose substrate.

23. A method for enzymatically modifying a cellulosic substrate, c h a r a c t e r - i z e d by contacting said substrate with an enzyme preparation according to any one of claims 20 to 22.

24. The method according to claim 23, wherein the cellulosic substrate is fibrous.

```
GATCTTCCATCTCGTGTCTTGCTTGTAACCATCGTGACCATGAAGGCAACTCTGGTTCTC  60
                                          MetLysAlaThrLeuValLeu   7

GGCTCCCTCATTGTAGGCGCCGTTTCCGCGTACAAGGCCACCACCACGCGCTACTACGAT 120
GlySerLeuIleValGlyAlaValSerAlaTyrLysAlaThrThrThrArgTyrTyrAsp  27

GGGCAGGAGGGTGCTTGCGGATGCGGCTCGAGCTCCGGCGCATTCCCGTGGCAGCTCGGC 180
GlyGlnGluGlyAlaCysGlyCysGlySerSerSerGlyAlaPheProTrpGlnLeuGly  47

ATCGGCAACGGAGTCTACACGGCTGCCGGCTCCCAGGCTCTCTTCGACACGGCCGGAGCT 240
IleGlyAsnGlyValTyrThrAlaAlaGlySerGlnAlaLeuPheAspThrAlaGlyAla  67

TCATGGTGCGGCGCCGGCTGCGGTAAATGCTACCAGCTCACCTCGACGGGCCAGGCGCCC 300
SerTrpCysGlyAlaGlyCysGlyLysCysTyrGlnLeuThrSerThrGlyGlnAlaPro  87

TGCTCCAGCTGCGGCACGGGCGGTGCTGCTGGCCAGAGCATCATCGTCATGGTGACCAAC 360
CysSerSerCysGlyThrGlyGlyAlaAlaGlyGlnSerIleIleValMetValThrAsn 107

CTGTGCCCGAACAATGGGAACGCGCAGTGGTGCCCGGTGGTCGGCGGCACCAACCAATAC 420
LeuCysProAsnAsnGlyAsnAlaGlnTrpCysProValValGlyGlyThrAsnGlnTyr 127

GGCTACAGCTACCATTTCGACATCATGGCGCAGAACGAGATCTTTGGAGACAATGTCGTC 480
GlyTyrSerTyrHisPheAspIleMetAlaGlnAsnGluIlePheGlyAspAsnValVal 147

GTCGACTTTGAGCCCATTGCTTGCCCCGGGCAGGCTGCCTCTGACTGGGGGACGTGCCTC 540
ValAspPheGluProIleAlaCysProGlyGlnAlaAlaSerAspTrpGlyThrCysLeu 167

TGCGTGGGACAGCAAGAGACGGATCCCACGCCCGTCCTCGGCAACGACACGGGCTCAACT 600
CysValGlyGlnGlnGluThrAspProThrProValLeuGlyAsnAspThrGlySerThr 187
                                        #     ↳ B

CCTCCCGGGAGCTCGCCGCCAGCGACATCGTCGAGTCCGCCGTCTGGCGGCGGCCAGCAG 660
ProProGlySerSerProProAlaThrSerSerSerProProSerGlyGlyGlyGlnGln 207
                                                B↲ A

ACGCTCTATGGCCAGTGTGGAGGTGCCGGCTGGACGGGACCTACGACGTGCCAGGCCCCA 720
ThrLeuTyrGlyGlnCysGlyGlyAlaGlyTrpThrGlyProThrThrCysGlnAlaPro 227

GGGACCTGCAAGGTTCAGAACCAGTGGTACTCCCAGTGTCTTCCTTGAGAAGGCCCAAGA 780
GlyThrCysLysValGlnAsnGlnTrpTyrSerGlnCysLeuPro                 242

TAGCCATGTCTCTCTAGCATTCTTCCGGCGTCAGTCTGATCTGCCTATTTAATCAGGTCA 840

GTCAATATGTATCCAGAGATAATAAATTATGTATATTATAGCAG(A)n              923
```

FIG. 1

SUBSTITUTE SHEET

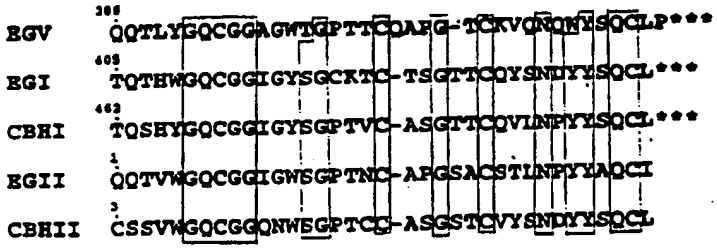

FIG. 2A

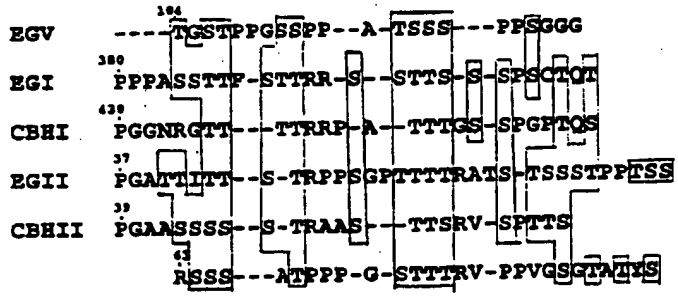

FIG. 2B

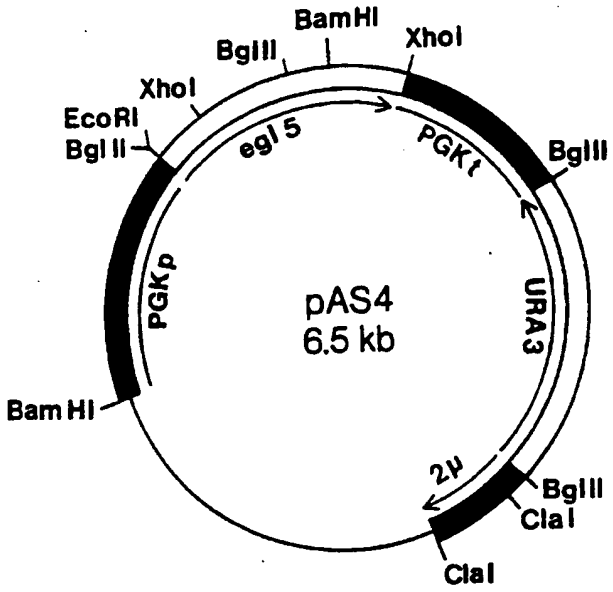

FIG. 3

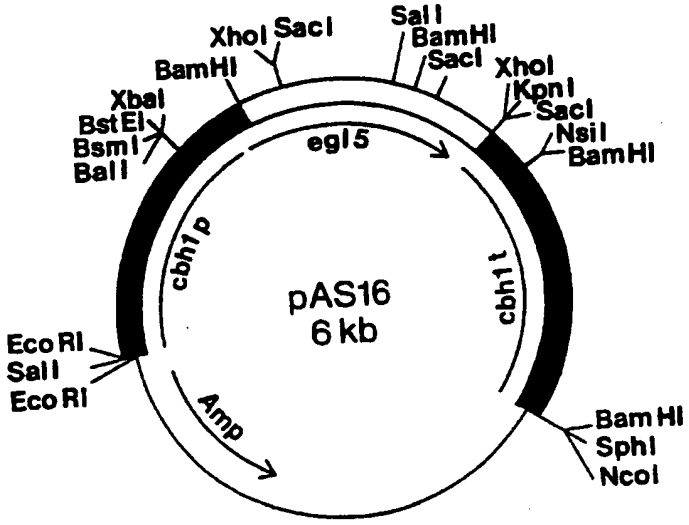

FIG. 4

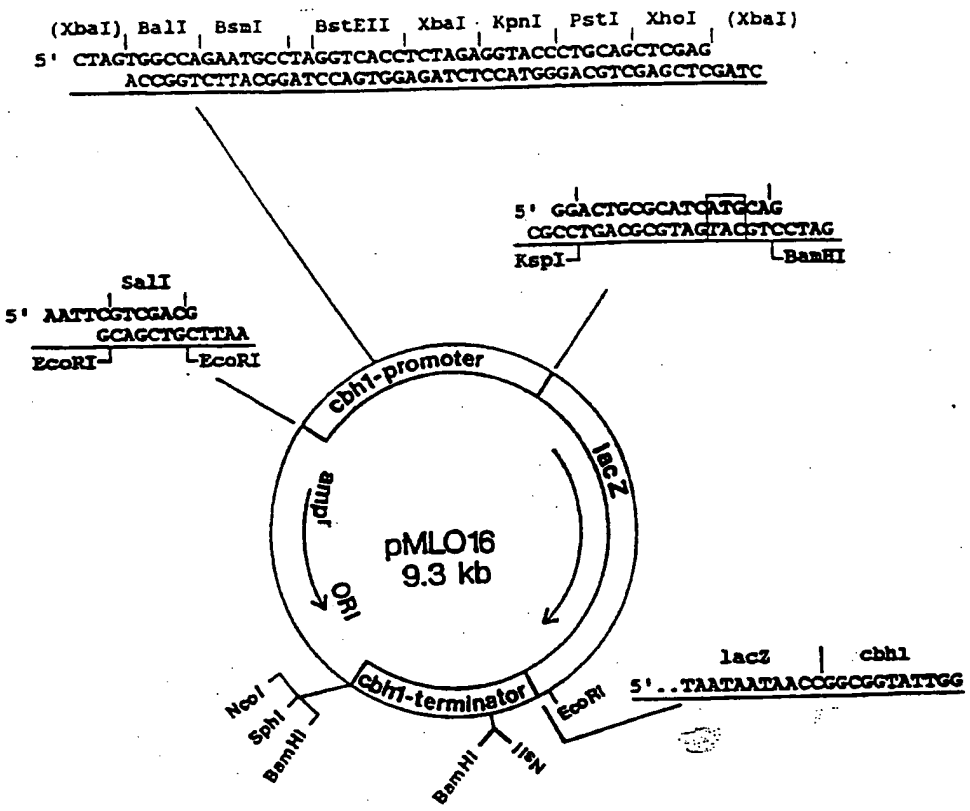

FIG. 5

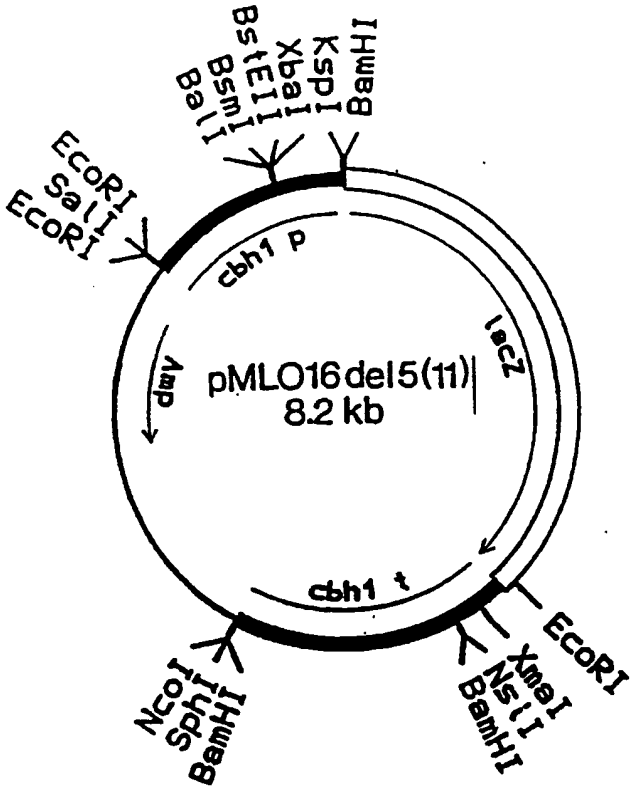

FIG. 6

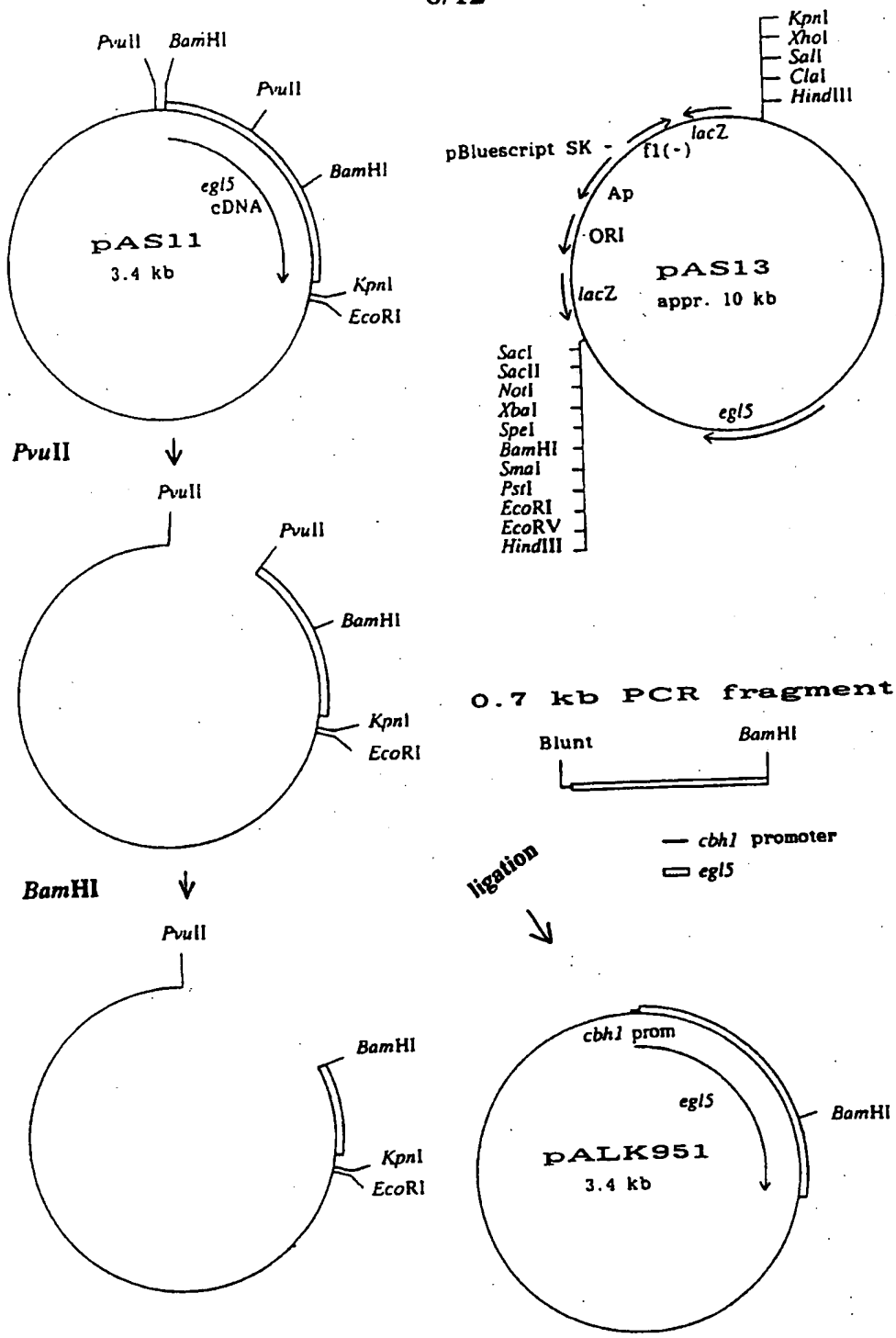

FIG. 7A

SUBSTITUTE SHEET

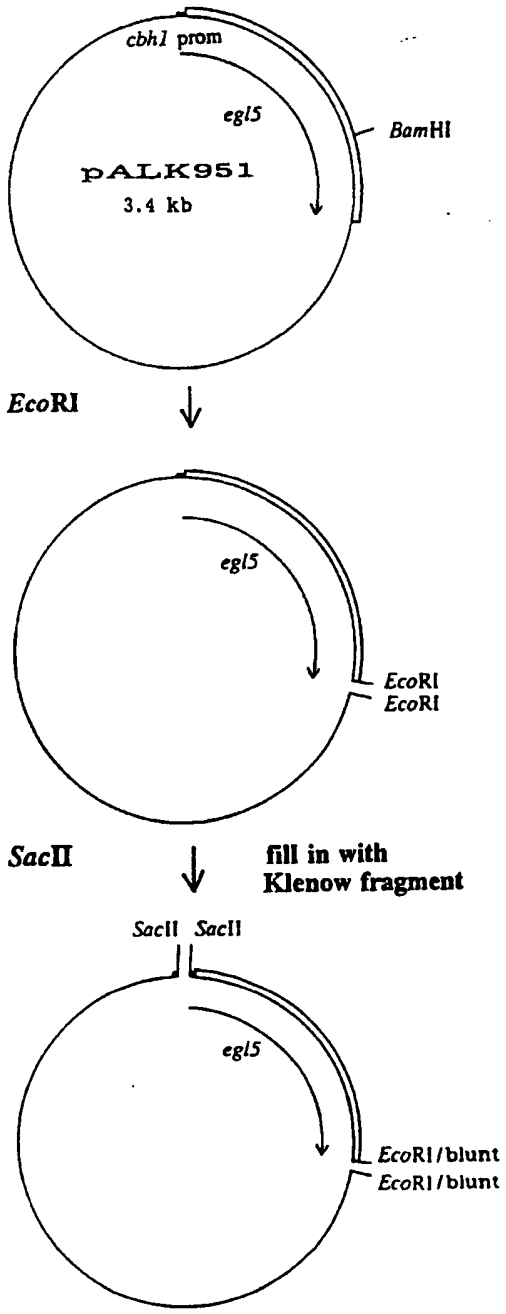

FIG. 7B

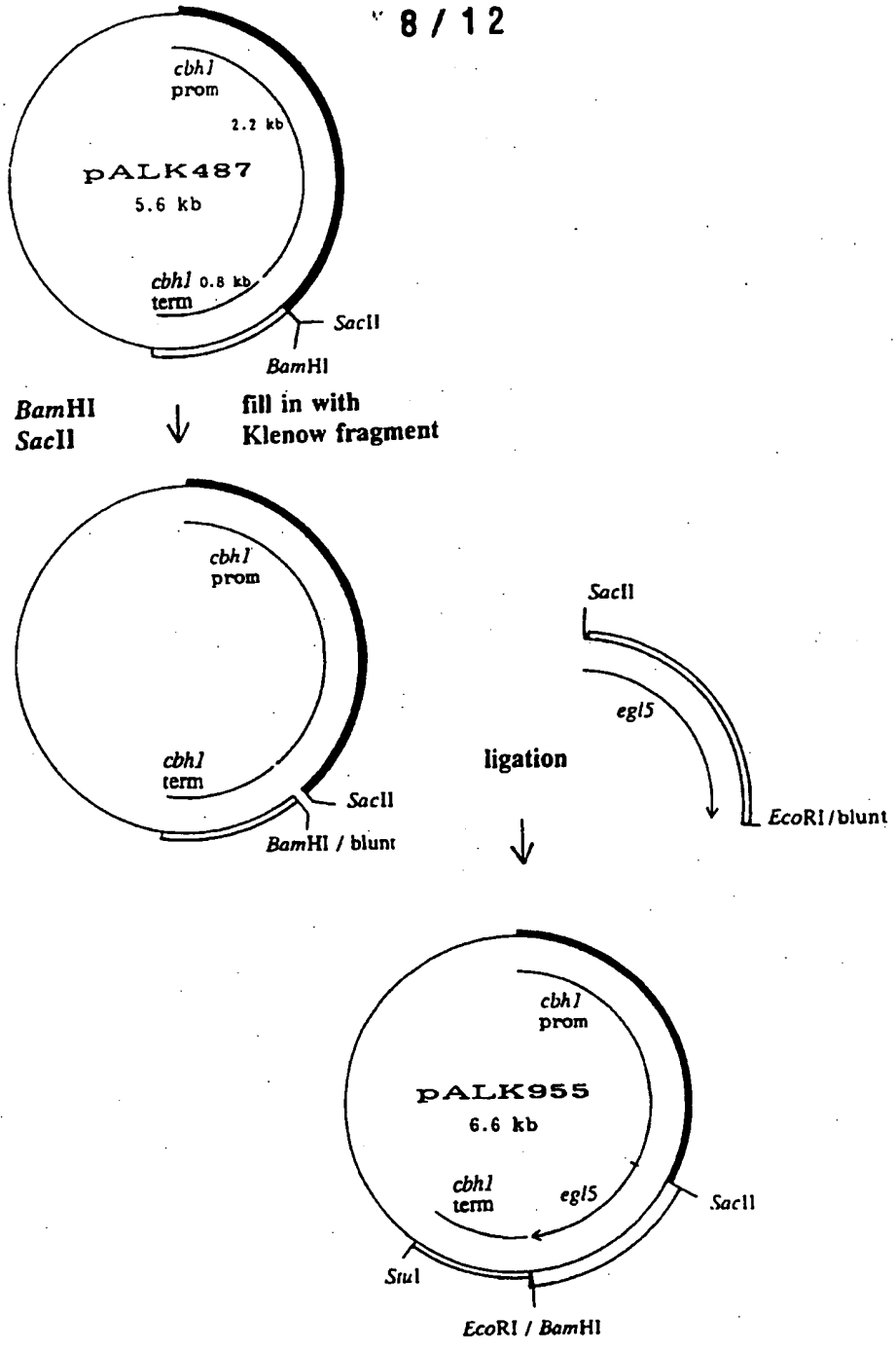

FIG. 7C

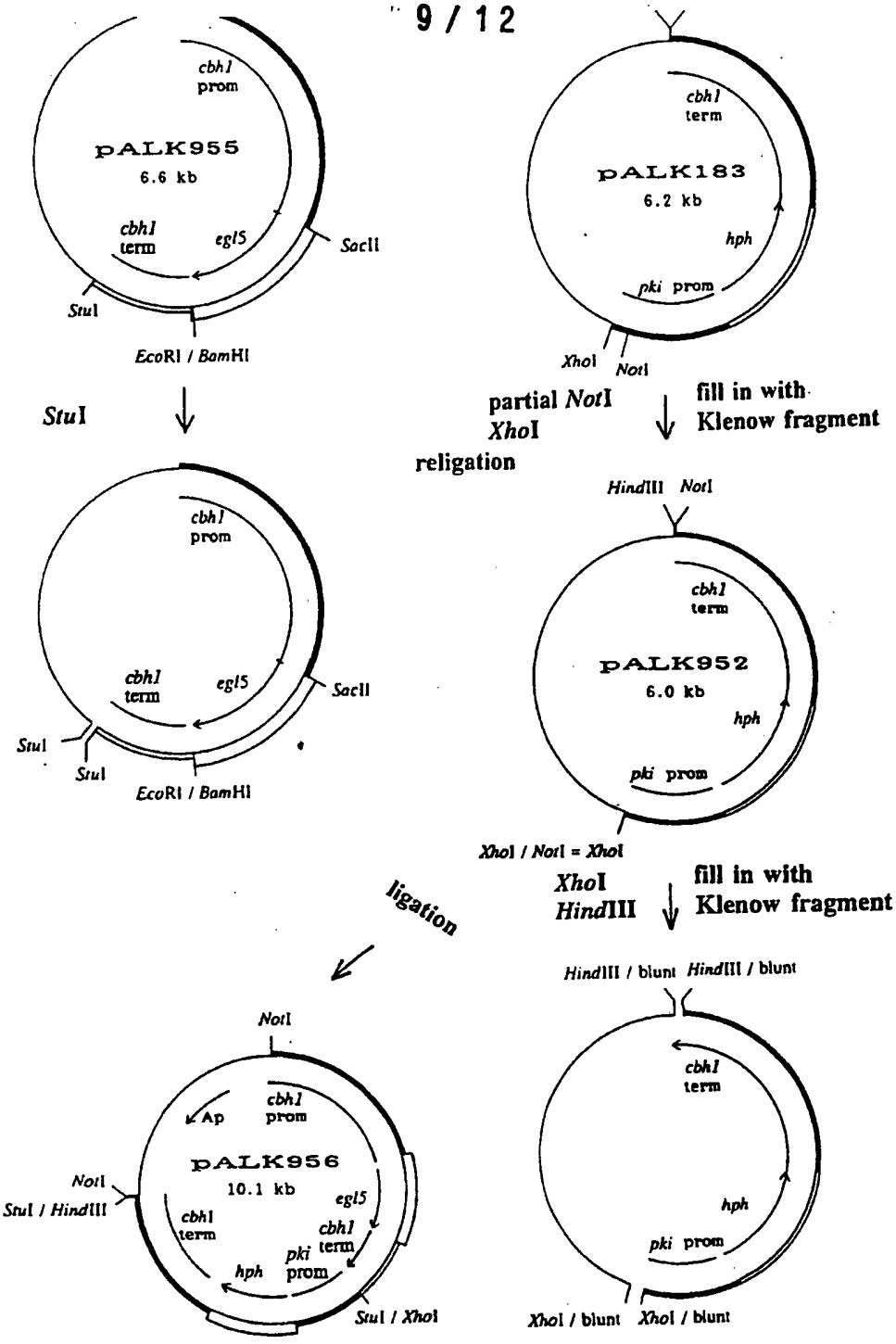

FIG. 7D

SUBSTITUTE SHEET

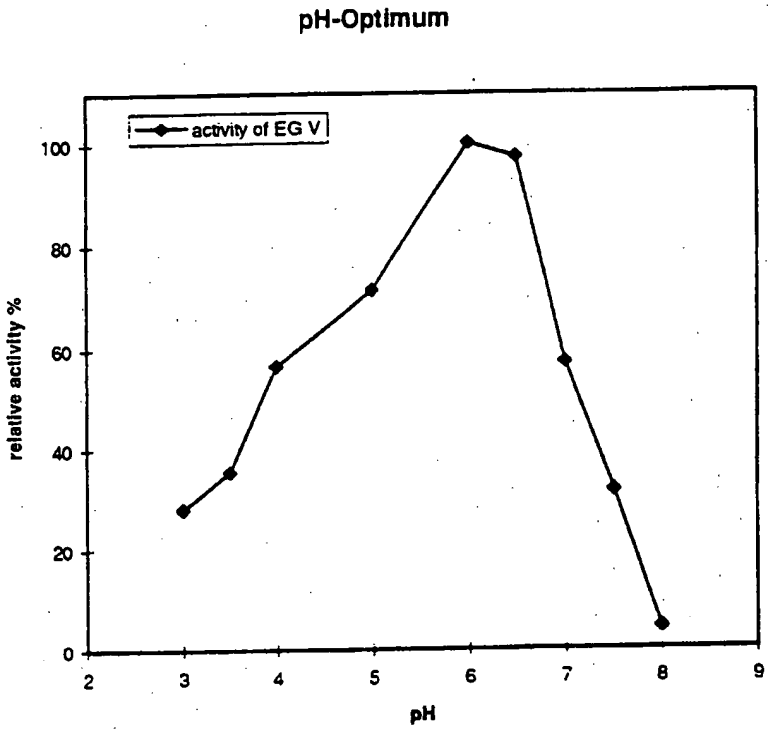

FIG. 8

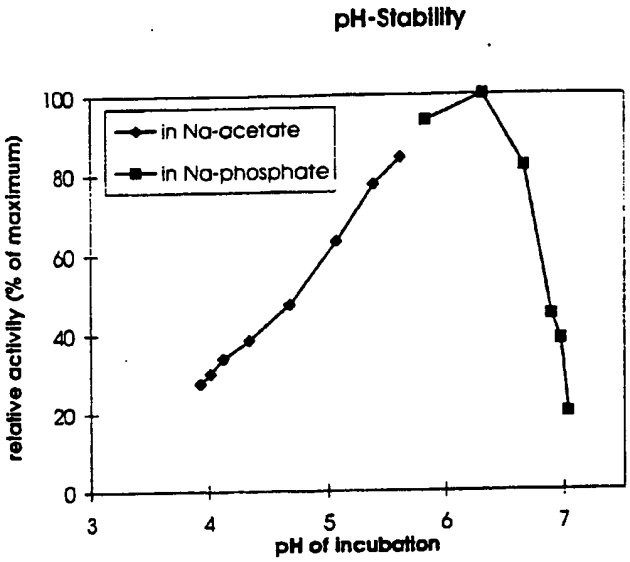

FIG. 9

```
GATCTTCCATCTCGTGTCTTGCTTGTAACCATCGTGACCATGAAGGCAACTCTGGTTCTC   60
                                                MetLysAlaThrLeuValLeu    7

GGCTCCCTCATTGTAGGCGCCGTTTCCGCGTACAAGGCCACCACCACGGCAAGTCTACAT  120
GlySerLeuIleValGlyAlaValSerAlaTyrLysAlaThrThrThr                23

GCTTCCAGGTCACAACGTCTGCTCAACAACCTCTAACCGAAAGGCCAGCGCTACTACGAT  180
            intron 1 (60 bp)                      ArgTyrTyrAsp    27

GGGCAGGAGGGTGCTTGCGGATGCGGCTCGAGCTCCGGCGCATTCCCGTGGCAGGTAAAC  240
GlyGlnGluGlyAlaCysGlyCysGlySerSerSerGlyAlaPheProTrpGln          45

ATTCGACCTTGTCTGGGGCAAGGGACTCGTCACTTACATCCTCTCTCTCTCTGCAGCTCG  300
                    intron 2 (62 bp)                      LeuG    46

GCATCGGCAACGGAGTCTACACGGCTGCCGGCTCCCAGGCTCTCTTCGACACGGCCGGAG  360
lyIleGlyAsnGlyValTyrThrAlaAlaGlySerGlnAlaLeuPheAspThrAlaGlyA   66

CTTCATGGTGCGGCGCCGGCTGCGGTAAATGCTACCAGCTCACCTCGACGGGCCAGGCGC  420
laSerTrpCysGlyAlaGlyCysGlyLysCysTyrGlnLeuThrSerThrGlyGlnAlaP   86

CCTGCTCCAGCTGCGGCACGGGCGGTGCTGCTGGCCAGAGCATCATCGTCATGGTGACCA  480
roCysSerSerCysGlyThrGlyGlyAlaAlaGlyGlnSerIleIleValMetValThrA  106

ACCTGTGCCCGAACAATGGGAACGCGCAGTGGTGCCCGGTGGTCGGCGGCACCAACCAAT  540
snLeuCysProAsnAsnGlyAsnAlaGlnTrpCysProValValGlyGlyThrAsnGlnT  126

ACGGCTACAGCTACCATTTCGACATCATGGCGCAGAACGAGATCTTTGGAGACAATGTCG  600
yrGlyTyrSerTyrHisPheAspIleMetAlaGlnAsnGluIlePheGlyAspAsnValV  146

TCGTCGACTTTGAGCCCATTGCTTGCCCCGGGCAGGCTGCCTCTGACTGGGGGACGTGCC  660
alValAspPheGluProIleAlaCysProGlyGlnAlaAlaSerAspTrpGlyThrCysL  166

TCTGCGTGGGACAGCAAGAGACGGATCCCACGCCCGTCCTCGGCAACGACACGGGCTCAA  720
euCysValGlyGlnGlnGluThrAspProThrProValLeuGlyAsnAspThrGlySerT  186

CTCCTCCCGGGAGCTCGCCGCCAGCGACATCGTCGAGTCCGCCGTCTGGCGGCGGCCAGC  780
hrProProGlySerSerProProAlaThrSerSerSerProProSerGlyGlyGlyGlnG  206

AGACGCTCTATGGCCAGTGTGGAGGTGCCGGCTGGACGGGACCTACGACGTGCCAGGCCC  840
lnThrLeuTyrGlyGlnCysGlyGlyAlaGlyTrpThrGlyProThrThrCysGlnAlaP  226

CAGGGACCTGCAAGGTTCAGAACCAGTGGTACTCCCAGTGTCTTCCTTGAGAAGGCCCAA  900
roGlyThrCysLysValGlnAsnGlnTrpTyrSerGlnCysLeuPro                242

GATAGCCATGTCTCTCTAGCATTCTTCCGGCGTCAGTCTGATCTGCCTATTTAATCAGGT  960

CAGTCAATATGTATCCAGAGATAATAAATTATGTATATTATAGCAG(A)39           1044
```

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI 94/00234

A. CLASSIFICATION OF SUBJECT MATTER

IPC5: C12N 9/42, C12N 15/56
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC5: C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC, MEDLINE, BIOSIS, STRAND

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TRENDS IN BIOTECHNOLOGY, Volume 5, Sept 1987, Jonathan Knowles et al, "Cellulase families and their genes", page 255 - page 261, the whole document, especially p. 257, c. 3, line 3-8 --	1-24
X	Chemical Abstracts, Volume 109, No 25, 19 December 1988 (19.12.88), (Columbus, Ohio, USA), Voragen, A. G. J. et al, "Cellulase of a mutant strain of Trichoderma viride", THE ABSTRACT No 225362p, Methods Enzymol. 1988, 160, 243-251 --	1-24

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"B" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 Sept 1994	20 -09- 1994
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Carl Olof Gustafsson Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI 94/00234

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO, A1, 9206184 (GENENCOR INTERNATIONAL, INC.), 16 April 1992 (16.04.92), see the whole document, especially Table I p. 32, line 21 - p. 33 line 13	1-24

--

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

27/08/94

International application No.
PCT/FI 94/00234

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A1- 9206184	16/04/92	CA-A- 2093423	01/12/92
		EP-A- 0586375	16/03/94

Form PCT/ISA/210 (patent family annex) (July 1992)